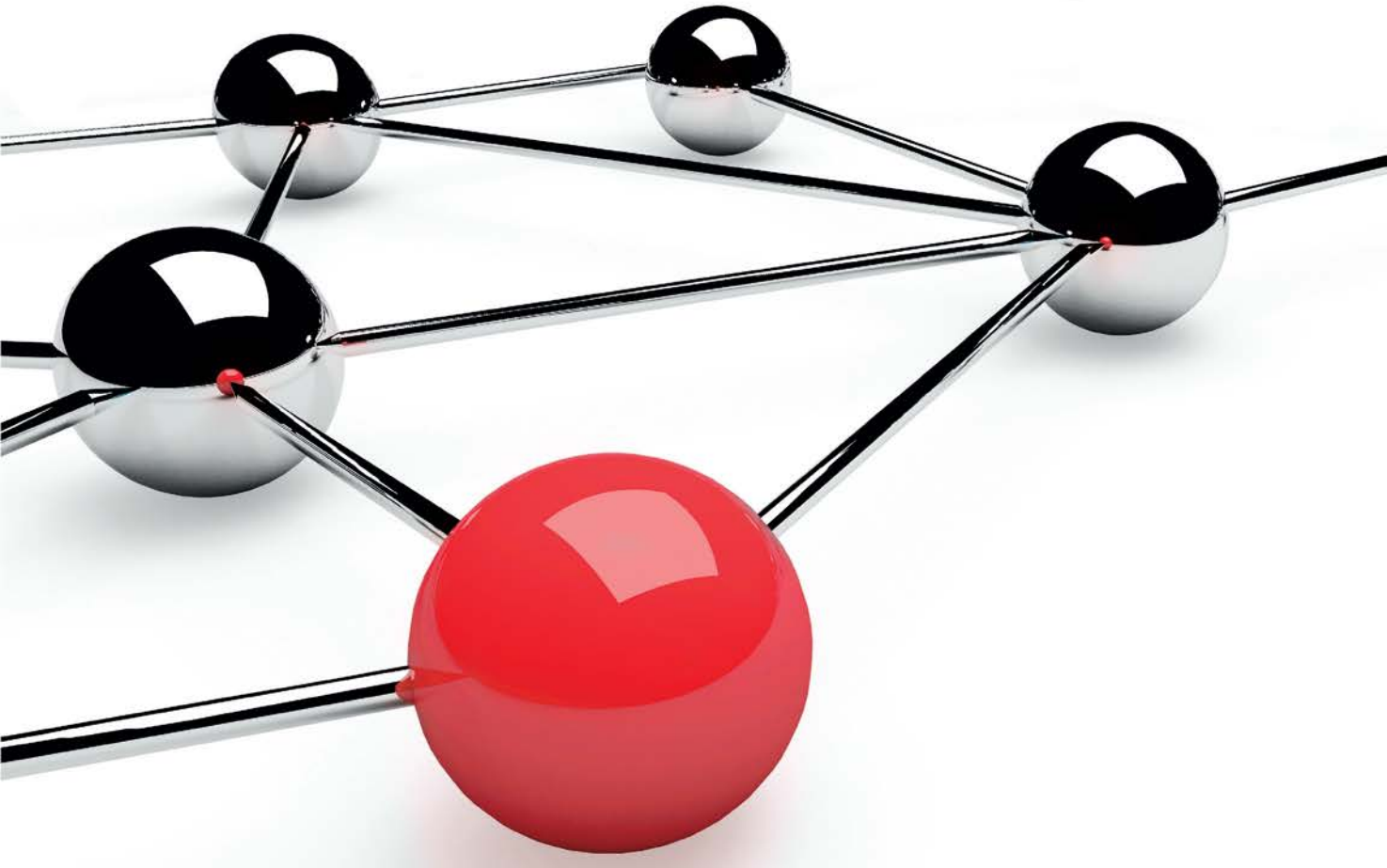


NINTH EDITION

ESSENTIALS OF ECONOMICS

JOHN SLOMAN
DEAN GARRATT

Pearson

ESSENTIALS OF ECONOMICS

Pearson

At Pearson, we have a simple mission: to help people make more of their lives through learning.

We combine innovative learning technology with trusted content and educational expertise to provide engaging and effective learning experiences that serve people wherever and whenever they are learning.

From classroom to boardroom, our curriculum materials, digital learning tools and testing programmes help to educate millions of people worldwide – more than any other private enterprise.

Every day our work helps learning flourish, and wherever learning flourishes, so do people.

To learn more, please visit us at **www.pearson.com**

ESSENTIALS OF ECONOMICS

NINTH EDITION

JOHN SLOMAN
The Economics Network, University of Bristol;
Visiting Professor, University of the West of England

DEAN GARRATT
Aston Business School

Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • São Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Pearson Education Limited
KAO Two
KAO Park
Harlow
CM17 9NA
United Kingdom
Tel: +44 (0)1279 623623
Web: www.pearson.com

Previously published 1998, 2001, 2004, 2007, 2010 (print), 2013, 2016, 2019 (print and electronic)
Ninth edition published 2023 (print and electronic)

ISBN: 978-1-292-44010-1 (print)
978-1-292-44015-6 (PDF)
978-1-292-44014-9 (ePub)

British Library Cataloguing-in-Publication Data
A catalogue record for the print edition is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Sloman, John, 1947- author. | Garratt, Dean, 1970- author.
Title: Essentials of economics / John Sloman, The Economics Network, University of Bristol, Visiting Professor, University of the West of England, Dean Garratt, Aston Business School.
Description: Ninth edition. | Harlow, England ; Hoboken, New Jersey : Pearson, 2023. | Includes bibliographical references and index. | Summary: "Welcome to this introduction to economics. Whether you are planning to study economics beyond this level, or whether this will be your only exposure to this fascinating subject, we hope that you will find the text enjoyable and that it will give you some insight into the economy in which you live and the economic forces that shape all our lives"– Provided by publisher.
Identifiers: LCCN 2022050399 (print) | LCCN 2022050400 (ebook) | ISBN 9781292440101 (paperback) | ISBN 9781292440156 (ebook) | ISBN 9781292440149 (epub)
Subjects: LCSH: Economics.
Classification: LCC HB171 .S59 2023 (print) | LCC HB171 (ebook) | DDC 330–dc23/eng/20221026
LC record available at https://lccn.loc.gov/2022050399
LC ebook record available at https://lccn.loc.gov/2022050400

10 9 8 7 6 5 4 3 2 1
27 26 25 24 23

All part and chapter opener images: John Sloman
Cover design by Kelly Miller
Cover image: © 3DProfi/Shutterstock

Print edition typeset in 10/12 pt Sabon MT Pro by Straive
Printed by Ashford Colour Press Ltd., Gosport

NOTE THAT ANY PAGE CROSS REFERENCES REFER TO THE PRINT EDITION

About the authors

John Sloman is Visiting Fellow at the University of Bristol and Associate of the Economics Network (www.economicsnetwork.ac.uk) a UK-wide organisation, where, until his retirement in 2012, he was Director. The Economics Network is based at the University of Bristol and provides a range of services designed to promote and share good practice in learning and teaching economics. The Network is supported by grants from the Royal Economic Society, the Scottish Economic Society and university economic departments and units from across the UK.

John is also visiting professor at the University of the West of England, Bristol, where, from 1992 to 1999, he was Head of School of Economics. He taught at UWE until 2007.

John has taught a range of courses, including economic principles on social science and business studies degrees, development economics, comparative economic systems, intermediate macroeconomics and managerial economics. He has also taught economics on various professional courses.

He is also the co-author with Dean Garratt and Jon Guest of *Economics* (Pearson Education, 10th edition 2018), with Dean Garratt, Jon Guest and Elizabeth Jones of *Economics for Business* (Pearson Education, 8th edition 2019) and with Elizabeth Jones of *Essential Economics for Business* (5th edition 2017). Translations or editions of the various books are available for a number of different countries with the help of co-authors around the world.

John is very interested in promoting new methods of teaching economics, including group exercises, experiments, role playing, computer-aided learning and use of audience response systems and podcasting in teaching. He has organised and spoken at conferences for both lecturers and students of economics throughout the UK and in many other countries.

As part of his work with the Economics Network he has contributed to its two sites for students and prospective students of economics: Studying Economics (www.studyingeconomics.ac.uk) and Why Study Economics? (www.whystudyeconomics.ac.uk)

From March to June 1997, John was a visiting lecturer at the University of Western Australia. In July and August 2000, he was again a visiting lecturer at the University of Western Australia and also at Murdoch University in Perth.

In 2007, John received a Lifetime Achievement Award as 'outstanding teacher and ambassador of economics' presented jointly by the Higher Education Academy, the Government Economic Service and the Scottish Economic Society.

Dean Garratt is a senior teaching fellow at Aston Business School. He joined Aston University in September 2018 having previously been a Principal Lecturer at Nottingham Business School. Dean teaches economics at a variety of levels, including modules in macroeconomics and economics for non-specialists. He is passionate about encouraging students to communicate economics more intuitively, to deepen their interest in economics and to apply economics to a range of issues.

Earlier in his career Dean worked as an economic assistant at both HM Treasury and at the Council of Mortgage Lenders. While at these institutions he was researching and briefing on a variety of issues relating to the household sector and to the housing and mortgage markets.

Dean is a Senior Fellow of the Higher Education Academy and an Associate of the Economics Network which aims to promote high-quality teaching practice. He has been involved in several projects promoting a problem-based learning (PBL) approach in the teaching of economics. In 2006, Dean was awarded the Outstanding Teaching Prize by the Economics Network. The award recognises exemplary teaching practice that deepens and inspires interest in economics. In 2013, he won the student-nominated Nottingham Business School teacher of the year award.

Dean has been an academic assessor for the Government Economic Service (GES) helping to assessed candidates at Economic Assessment Centres (EACs). In this role he assessed candidates looking to join the GES, the UK's largest employer of professional economists.

Dean has run sessions on HM Treasury's Graduate Development Programme (GDP). These sessions covered principles in policy making, applying economics principles and ideas to analyse policy issues and contemporary developments in macroeconomics.

Outside of work, Dean is an avid watcher of many sports. Having been born in Leicester, he is a season ticket holder at both Leicester City Football Club and Leicestershire County Cricket Club.

Brief contents

Detailed contents

Part C MACROECONOMICS

Preface

TO THE STUDENT

Welcome to this introduction to economics. Whether you are planning to study economics beyond this level, or whether this will be your only exposure to this fascinating subject, we hope that you will find the text enjoyable and that it will give you some insight into the economy in which you live and the economic forces that shape all our lives.

Although you have probably never studied the subject before, you will almost certainly know quite a lot of economics already. After all, you make economic decisions virtually every day of your life. Every time you go shopping, you are acting as an 'economist': deciding what to buy with your limited amount of money. And in a period of rising prices, such as in 2022, these decisions become more important. But it is not just with decisions about buying that we act as economists. How much to work (something that students are increasingly forced to do nowadays), how much to study, even how much time to devote to various activities during the course of the day, are all, in a way, *economic* choices.

To satisfy us as consumers, goods and services have to be produced. We will therefore study the behaviour of firms and what governs the decisions that they make. How will the decisions of big businesses differ from those of small firms? How will the degree of competition affect the extent to which we gain or lose from the activities of firms?

In analysing economic choices we look at some of the big economic issues that face us all as members of society in the twenty-first century. Despite huge advances in technology, and despite the comfortable lives led by many people in the industrialised world, we continue to suffer from volatile economic growth, industrial change and unemployment and all the insecurity that these bring. Unexpected shocks, such as the COVID-19 pandemic, can have a major effect on livelihoods and jobs. We continue to witness poverty and inequality, and in many countries the gap between rich and poor has actually grown wider; our environment is polluted and the planet is warming; our growing affluence as consumers is increasingly bought at the expense of longer hours at work and growing levels of stress.

We live in a highly interdependent world where actions have implications elsewhere. The Russian invasion of Ukraine directly caused rises in gas, oil and grain prices and drove up the cost of living, hitting poor people especially hard as they spend a large proportion of their income on energy and food.

So, what can be done about these problems? This text seeks not only to analyse these problems but also to examine the sorts of policies that governments might pursue in their attempt to address them.

The text is designed with one overriding aim: to make this exciting and highly relevant subject as clear to understand as possible. To this end, the text has a number of important features:

- A direct and straightforward written style; short paragraphs to aid rapid comprehension. The aim all the time is to provide maximum clarity.
- A careful use of colour to guide you through the text and make the structure easy to follow.
- Key ideas highlighted and explained where they first appear. These ideas are key elements in the economist's 'toolkit'. Whenever they recur later in the text, an icon appears in the margin and you are referred back to the page where they are defined and explained. All the key ideas are gathered together at the beginning of the Glossary.
- Some of the key ideas are particularly important in affecting the way we see the world: they help us think like economists. We call these 'threshold concepts' and there are 15 of these.
- Clear chapter-opening pages, which set the scene for the chapter. They also highlight the issues that will be covered in the chapter and can thus be seen as 'learning objectives'.
- Summaries at the end of each section (rather than each chapter). This provides a very useful means of revising and checking your understanding as you progress.
- Definitions of all technical terms given at the foot of the page where the term is first used. The term itself is highlighted in the text.
- 'Pause for thought' questions integrated in the text. These are designed to help you reflect on what you have just read and to check on your understanding. Answers to all 'pause for thought' questions are given on the student free-access companion website, which, for the rest of the text, we refer to simply as the 'student website'.
- A comprehensive index. This enables you to look up a concept or topic as required and to see it used in context.
- An alphabetical glossary at the end of the text. This gathers together all the defined terms.
- Plentiful use of up-to-date examples to illustrate the arguments. This helps to bring the subject alive and puts it in context.

- Review questions at the end of each chapter for either individual or class use.
- Answers to all odd-numbered questions are given on the student website. These questions will be helpful for self-testing, while the even-numbered ones can be used for class testing.
- Many boxes (typically four to six per chapter) providing case studies, news items, applications, or elaborations of the text. The boxes are of two types: Case Studies and Applications; and Exploring Economics. Each box contains questions allowing students to assess their own understanding and each box contains an activity designed to develop important skills around research, data analysis and the communication of economic ideas and principles. These skills are not only of use to you at university but also in the world of work. They are frequently identified by employers as being especially valuable. Hence, by undertaking the activities in the boxes you help increase your employability.
- A comprehensive set of web references at the end of each of the four parts of the text. Each reference is numbered to match those in the Web Appendix at the end of the text. You can easily access any of these sites from this book's own website (at http://www.pearsoned.co.uk/sloman). When you enter the site, click on **Hot Links.** You will find all the sites from the Web Appendix listed. Click on the one you want and the 'hot link' will take you straight to it.
- Appendices for most chapters appear on the student website. These Web Appendices take the argument further than in the text and look at some more advanced theories. While none of these is necessary for studying this text, and many courses will not refer to them, they provide the necessary additional material for more advanced courses that still require a short textbook.

Good luck with your studies – and have fun. Perhaps this will be just the beginning for you of a lifelong interest in economic issues and the economy.

TO LECTURERS AND TUTORS

This ninth edition of *Essentials of Economics* is an abridged version of *Economics,* 11th edition (John Sloman, Dean Garratt and Jon Guest). Some passages have been directly transcribed, while others have been extensively rewritten in order to provide a consistent coverage of the 'essentials' of economics.

The text is designed specifically for one-semester courses in introductory economics. There are 15 chapters (1 introductory, 7 micro, 5 macro and 2 international), each providing about a week's worth of reading. The text is also ideal for year-long courses that are designed for those not going on to specialise in economics, or where economics is only a subsidiary component.

Naturally, in a one-semester course, or in courses for non-specialists, tutors cannot hope to cover all the principles of economics. Thus some things have had to go. The text does not cover indifference curves or isoquants. The analysis of costs is developed with only an informal reference to production functions. Distribution theory is confined to the determination of wage rates. In macroeconomics, *IS*/*LM* and *IS*/*MP* analysis have been left out, as have some of the more advanced debates in monetary and exchange rate theory. In addition, many passages have been simplified to reflect the nature of courses on which the text is likely to be used. The result is a text that is approximately half the length of *Economics,* 11th edition.

Suggestions for longer or more advanced courses

If you want to use this text on more rigorous courses, most chapters have one or more Web Appendices. These introduce students to more advanced models, such as indifference analysis, isoquant analysis, general equilibrium in both a closed and an open economy, *IS*/*LM, IS*/*MP,* the full money multiplier, and trade creation and diversion. You can use any or all of them to fit your course.

The text is also ideal for the economics AS/A2 syllabuses of the various boards.

The text as also highly suitable for courses, such as HND, where the economic environment component is part of a larger module.

Extensive revision within the existing structure

To bring economics alive and show how the subject relates to real-world issues, the ninth edition of *Essentials of Economics* contains a great deal of applied material. Consequently, there have been considerable revisions from the previous editions to reflect contemporary issues, debates and policy interventions. Specifically, you will find that:

- Many of the boxes are new or extensively revised.
- There are many new examples given in the text.
- All tables and charts have been updated, as have factual references in the text.
- Economic analysis and debate has been strengthened and revised at various points in the text in light of economic events and developments in economic thinking.
- Building on the revisions in previous editions there is considerable coverage of behavioural economics.

- The climate emergency is considered in many places throughout the text, with consideration of the economic causes and government policies to tackle global warming.
- There is also coverage throughout of the impact of the COVID-19 pandemic and the various measures taken by governments to limit the economic damage.
- There is analysis of the impact on inflation of both supply problems in the aftermath of COVID and the Russian invasion of Ukraine. The reaction of governments and central banks is also considered.
- All policy sections have been thoroughly revised to reflect the changes that have taken place since the last edition. This includes an analysis in several parts of the text of the implications of the UK's exit from the EU and of policies being pursued in various parts of the world that restrict trade.
- Most importantly, every part of the text has been carefully considered, and if necessary redrafted, to ensure both maximum clarity and contemporary relevance.

The text also contains 36 'key ideas' and these are highlighted and explained when they first appear. These fundamental concepts provide a 'toolkit' for students. Students can see them recurring throughout the text, and an icon appears in the margin to refer back to the page where the idea first appears. Showing how these ideas can be used in a variety of contexts helps students to relate the different parts of the subject to each other. Fifteen of these concepts are given the special status of 'Threshold Concepts'. Understanding and being able to use these concepts, such as opportunity cost, help students to 'think like an economist'. Each of these concepts is explained in detail in MyLab Economics and on the student companion website.

But, despite considerable updating, the book retains the same structure as the previous edition. This should make the transition to the ninth edition straightforward as it removes the need to update references to chapters, sections and pages in the book.

We hope that your students will find this an exciting and interesting text that is relevant to today's issues.

SUPPLEMENTS

MyLab Economics for students

MyLab Economics provides a comprehensive set of online tests, homework and revision exercises. If you have purchased this text as part of a pack, then you can gain access to MyLab by following the instructions to register the access code. If you've purchased this text on its own, then you can purchase access online at www.myeconlab.com

MyLab Economics provides a variety of tools to enable students to access their own learning, including exercises, quizzes and tests, arranged chapter by chapter. There are many new questions in this edition and each question has been carefully considered to reflect the learning objectives of the chapter. A personalised Study Plan identifies areas to concentrate on to improve grades, and specific tools are provided to each student to direct their studies in a more efficient way.

Student website

In addition to the materials on MyLab Economics, there is an open-access companion website for students with a large range of other resources, including:

- Animations of key models with audio explanations. These 'audio animations' can be watched online or downloaded to a computer, MP4 player, smart phone, etc.
- Links to the Sloman Economics News site with news items added several times each month, with introductions, links to newspaper and other articles and to relevant data, questions for use in class or for private study, and references to chapters in the text. You can search the extensive archive by chapter or keyword.
- More than 200 case studies with questions for self-study, ordered chapter by chapter and referred to in the text.
- A set of Web appendices which explore economic theory further than in the text and are suitable for courses with more advanced sections or where students want to study the subject in greater depth.
- An updated list of over 280 hot links to sites relevant to economics. These are referred to in the book's Websites Appendix and at the end of each of the four Parts of the text.
- Answers to all Pause for thought questions.
- Answers to end-of-chapter questions.
- Threshold Concepts. A detailed description of each of the 15 Threshold Concepts, showing how understanding them and being able to apply them in a variety of contexts helps you to think like an economist.

Note that Sloman Economics News and hotlinks can also be accessed directly from http://pearsonblog.campaignserver.co.uk/.

MyLab Economics for lecturers and tutors

You can register online at www.myeconlab.com to use MyLab Economics, which is a complete virtual learning environment for your course or embedded into Blackboard, Moodle or your university's own online learning platform. You can customise its look and feel and its availability to students. You can use it to provide support to your students in the following ways:

- MyLab's gradebook automatically records each student's time spent and performance on the tests and Study Plan. It also generates reports you can use to monitor your students' progress.

- You can use MyLab to build your own test, quizzes and homework assignments from the question base provided to set your own students' assessment.
- A select number of questions are generated algorithmically so they use different values each time they are used.
- You can create your own exercises by using the econ exercise builder.

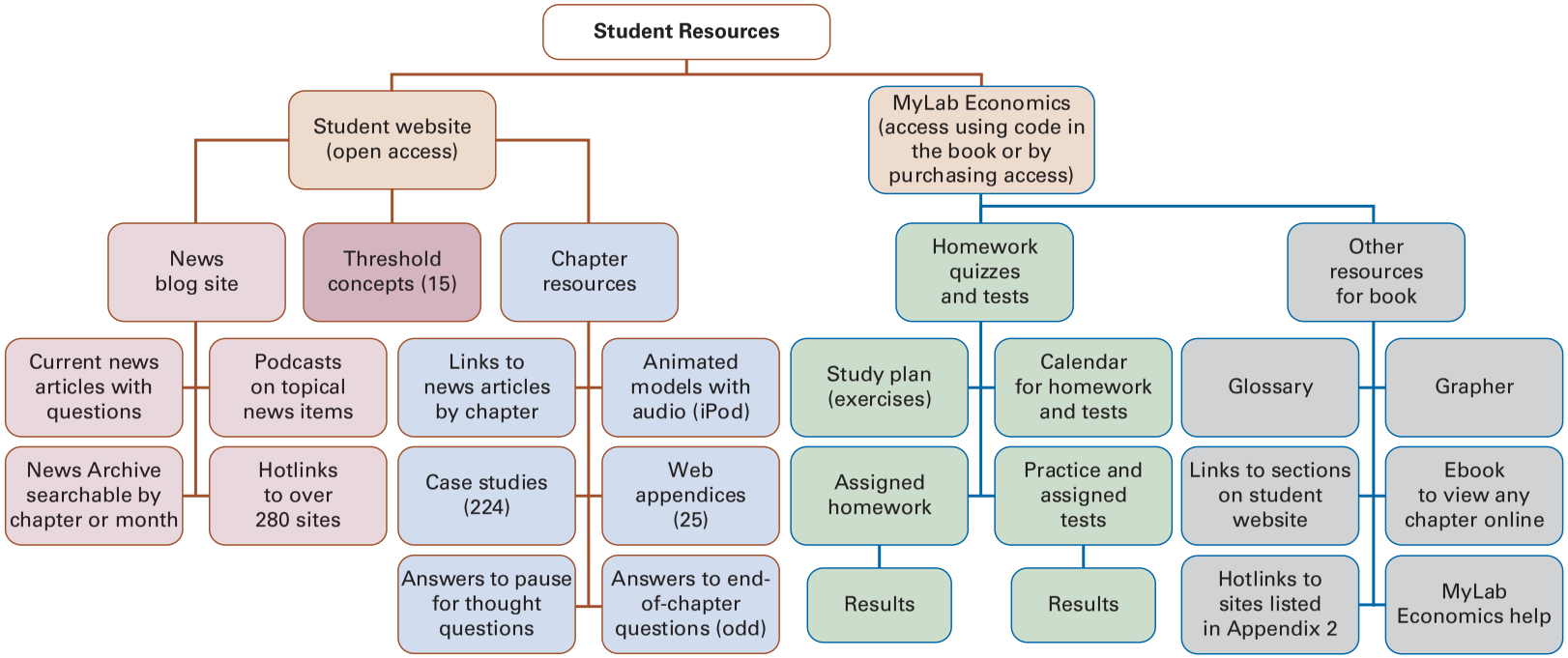
Contact your local Pearson representative for more details and support.

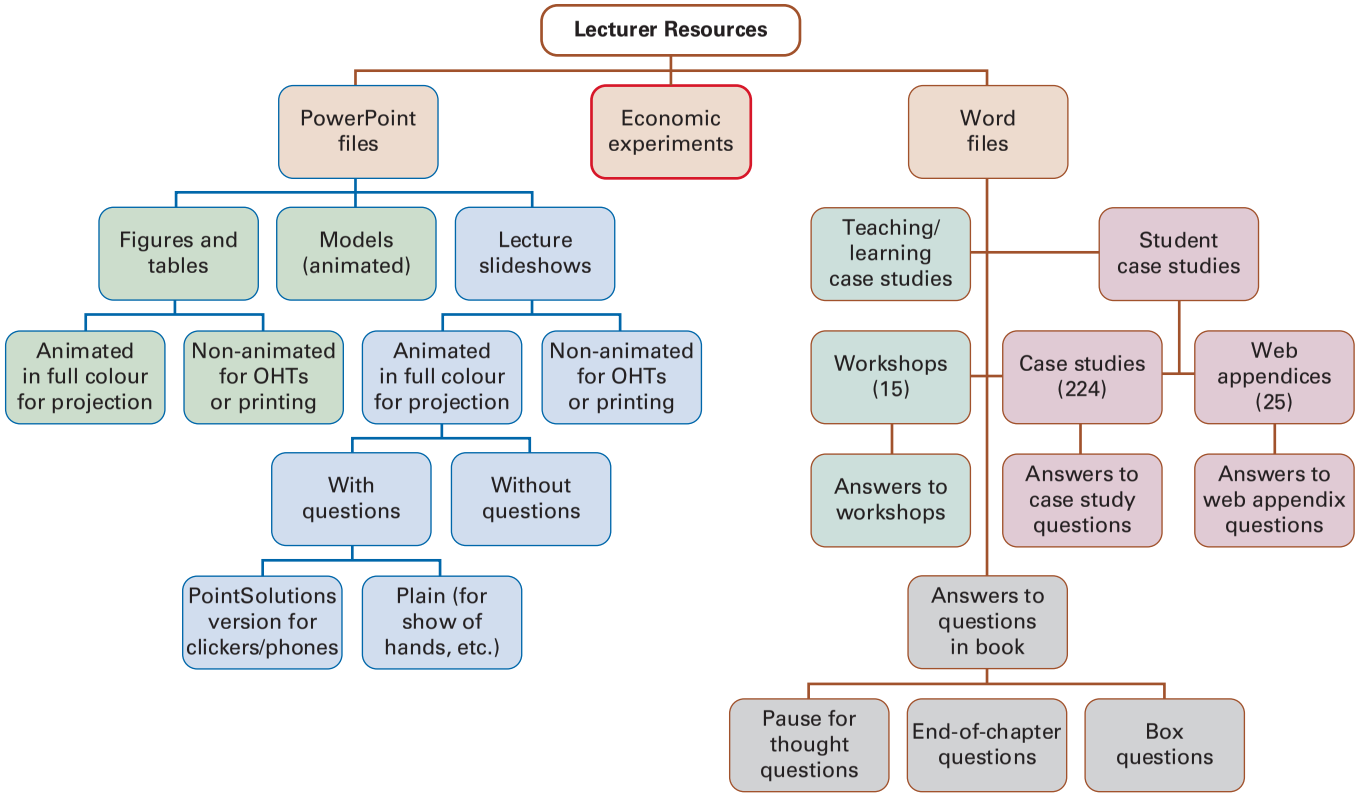
Additional resources for lecturers and tutors

There are many additional resources for lecturers and tutors that can be downloaded from the lecturer section of MyLab or from the Lecturer Resources section of the book's website at www.pearsoned.co.uk/sloman. These have been thoroughly revised for the ninth edition. These include:

- PowerPoint® slide shows in full colour for use with a data projector in lectures and classes. These can also be made available to students by loading them on to a local network. There are several types of slideshows:
  - *All figures from the text and most of the tables*. Each figure is built up in a logical sequence, thereby allowing tutors to show them in lectures in an animated form. There is also a non-animated version suitable for printing or for display on an OHP or visualiser.
  - *A range of models*. There are 26 files, each containing one of the key models from the text, developed in an animated sequence of between 20 to 80 screens.
  - *Customisable lecture slideshows*. These are a series of bullet-point lecture plans. There is one for each chapter of the text. Each one can be easily edited, with points added, deleted or moved, so as to suit particular lectures. A consistent use of colour is made to show how the points tie together. It is not intended that all the material is covered in a single lecture; you can break at any point. It's just convenient to organise them by chapter. They come in various versions:
  - Lecture slideshows with integrated diagrams. These lecture plans include animated diagrams, charts and tables at the appropriate points.
  - Lecture plans with integrated diagrams and questions. These are like the above but also include multiple-choice questions, allowing lectures to become more interactive. They can be used with or without an audience response system (ARS). A special ARS version is available for TurningPoint® and is ready to use with appropriate 'clickers' or with smartphones, tablets or laptops.
  - Lecture plans without the diagrams. These allow you to construct your own diagrams on the blackboard, whiteboard or visualiser or use pre-prepared ones on a visualiser or OHP.
- Case studies. These, also available in MyLab economics and on the student website, can be reproduced and used for classroom exercises or for student assignments. Answers are also provided (not available on the student site).
- Workshops. There are 15 of these – one for each chapter. They are in Word® and can be reproduced for use with large groups (up to 200 students) in a lecture theatre or large classroom. In A-level classes, they can be used as worksheets, either for use in class or for homework. Suggestions for use are given in an accompanying file. Answers to all workshops are given in separate Word files.
- Teaching/learning case studies. There are 20 of these. They examine various approaches to teaching introductory economics and ways to improve student learning.
- Answers to all end-of-chapter questions, pause for thought questions, questions in boxes, questions in Web Cases and Web Appendices and to the 15 workshops. They have been completely revised with new hyperlinks where appropriate.

The following two pages show in diagrammatic form all the student and lecturer resources.


Student Resources
Student website (open access)
MyLab Economics (access using code in the book or by purchasing access)
News blog site
Threshold concepts (15)
Chapter resources
Homework quizzes and tests
Other resources for book
Current news articles with questions
Podcasts on topical news items
Links to news articles by chapter
Animated models with audio (iPod)
Study plan (exercises)
Calendar for homework and tests
Glossary
Grapher
News Archive searchable by chapter or month
Hotlinks to over 280 sites
Case studies (224)
Web appendices (25)
Assigned homework
Practice and assigned tests
Links to sections on student website
Ebook to view any chapter online
Answers to pause for thought questions
Answers to end-of-chapter questions (odd)
Results
Results
Hotlinks to sites listed in Appendix 2
MyLab Economics help


Lecturer Resources
PowerPoint files
Economic experiments
Word files
Figures and tables
Models (animated)
Lecture slideshows
Teaching/ learning case studies
Student case studies
Animated in full colour for projection
Non-animated for OHTs or printing
Animated in full colour for projection
Non-animated for OHTs or printing
Workshops (15)
Case studies (224)
Web appendices (25)
With questions
Without questions
Answers to workshops
Answers to case study questions
Answers to web appendix questions
PointSolutions version for clickers/phones
Plain (for show of hands, etc.)
Answers to questions in book
Pause for thought questions
End-of-chapter questions
Box questions

ACKNOWLEDGEMENTS

As with previous editions, we owe a debt to various people. A special thanks to Peter Smith from the University of Southampton for authoring the MyLab Economics questions and tests. Thanks to the team at Pearson Education, and especially to Catherine Yates, who has been a tremendous help and support at every stage of revising the text. Thanks to Mel Carter, Jodie Mardell-Lines and Vivek Khandelwal for all the work they have put in to producing the text and its supplements. Thanks too to the many users of the text who have given us feedback. We always value their comments.

John: As always, I owe a huge debt to my family, and especially my wife and soulmate Alison, whose love and support have made this and previous editions possible. And many thanks once again to Dean, whose ideas and enthusiasm have been fantastic. It's been great to work together.

Dean: A special thank you must go to Patricia. She continues to be my rock and remains steadfastly supportive. I would like to thank my parents for all their love and continued support. Finally, thanks to John for again inviting me to work on *Essentials of Economics* and, not least, for his abundant patience.

PUBLISHER'S ACKNOWLEDGEMENTS

Text credits

11 European Commission: Based on data in AMECO Database, European Commission, DG Business, Economy, Euro (May 2022); **18 Office for National Statistics:** Consumer Price Inflation time series dataset and UK House Price Index: reports (Office for National Statistics); **19 Office for National Statistics:** Consumer Price Inflation time series dataset and UK House Price Index: reports (Office for National Statistics); **23 International Monetary Fund:** Data drawn from World Economic Outlook Database (IMF, April 2022) and various forecasts; **38 Office for National Statistics:** Based on data from Consumer Price Inflation time series dataset and UK House Price Index: reports (Office for National Statistics); **40 Office for National Statistics:** Based on data from Consumer Price Inflation time series dataset, series CHAW (ONS) and various (2022); **42 The World Bank Group:** Based on data from World Bank Commodity Price Data (The Pink Sheet), (Commodity Markets, World Bank, 2022); **97 Macmillan Publishers:** T.R. Malthus, First Essay on Population (Macmillan,1926), pp. 13–14.; **109 European Commission:** C. F. Pratten, 'A survey of the economies of scale', in Research into the 'Costs of Non-Europe', Volume 2 (Commission of the European Communities, Luxembourg, 1988).; **109 European Commission:** European Commission/ Economists Advisory Group Ltd, 'Economies of scale', The Single Market Review, Sub-series V, Volume 4 (Commission of the European Communities, Luxembourg, 1997).; **128 Office for National Statistics:** Based on series J4MC from Time Series Data (National Statistics).; **142 The World Bank Group:** Nominal oil price data from World Commodity Price Data (The Pink Sheet), Commodity Markets (World Bank); **142 Organisation for Economic Co-operation and Development:** Price Index from Data Extracts (OECD); **164 Office for National Statistics:** Based on data in series MGRZ, YCBH, MGRQ, YCBW and EMP17 (ONS); **173 Office for National Statistics:** Based on data from The effects of taxes and benefits on UK household income UK, 2020/21, Reference Table 2, (ONS, 28 March 2022); **175 Office for National Statistics:** The Effects of Taxes and Benefits on Household Income, UK, 2020/21, Reference Table 9 (ONS, 28 March 2022); **175 Organisation for Economic Co-operation and Development:** Based on data in Income Inequality, OECD dataset (accessed 21 April 2021), https://data.oecd.org/inequality/income-inequality.htm; **176 Office for National Statistics:** Based on data inThe Effects of Taxes and Benefits on Household Income, UK, 2020/21, Reference Table 28 (ONS, 28 March 2022); **177 Office for National Statistics:** Based on data from The effects of taxes and benefits on household income, UK, 2020/21, Reference Table 13, (ONS, 28 March 2022); **177 Office for National Statistics:** Based on data in the Annual Survey of Hours and Earnings, Table 14.2a, (National Statistics, October 2021); **178 Office for National Statistics:** Based on data in Gender pay gap in the UK: 2021 (ONS, October 2021); **179 Office for National Statistics:** Earnings and Hours Worked, occupation by four-digit SOC: ASHE Table 14.6a (ONS, October 2021); **180 Office for National Statistics:** Household total wealth in Great Britain: April 2018 to March 2020 (ONS, January 2022); **181 Office for National Statistics:** Based on data in Household Disposable Income and Inequality, Table 11 (National Statistics, January 2018); **181 Office for National Statistics:** Total Wealth: Wealth in Great Britain, July 2006 to June 2016, Table 2.5 (National Statistics, February 2018); **184 INTERNATIONAL MONETARY FUND:** 'Tackling Inequality, October 2017', Fiscal Monitor, IMF (October 2017).; **187 European Commission:** Based on data in AMECO Database, European Commission (to 1980), https://ec.europa.eu/info/business-economy-euro/indicators-statistics/economic-databases/macro-economic-database-ameco/ameco-database_en; **189 The National Archives:** 'Universal credit roll out delayed again – to 2024', BBC News (3 February 2020); **220 Office for National Statistics:** Based on data from Environmental taxes in the United Kingdom (Office for National Statistics).; **221 Organisation for Economic Co-operation and Development:** Based on data in Environmentally related tax revenue (OECD, 2022); **234 International Monetary Fund:** Based on data in AMECO Database

(European Commission) up to 1980; and World Economic Outlook (IMF, April 2022); **235 International Monetary Fund:** Based on data in data in AMECO Database (European Commission) up to 1980; and World Economic Outlook (IMF, April 2022); **236 European Commission:** Based on data in data in AMECO Database (European Commission) up to 1980; and World Economic Outlook (IMF, April 2022); **236 European Commission:** Based on data in AMECO Database (European Commission) up to 1980; and World Economic Outlook (IMF, April 2022); **239 Office for National Statistics:** Based on series CGDA and YBHA (ONS); **242 European Commission:** Based on data from AMECO database (European Commission, DGECFIN, May 2022) and various forecasts; **243 European Commission:** Based on data in AMECO Database, European Commission, DGECFIN; **244 United Nations:** National Accounts Main Aggregates Database (United Nations Statistics Division), https://unstats.un.org/unsd/snaama/; **259 Office for National Statistics:** Based on series KGZ7, KG7T and IHYR (ONS, 2022); **260 Office for National Statistics:** Based on data from The UK National balance sheet estimates: 2021 and series YBHA and HABN (National Statistics, 2022); **264 European Commission:** Based on data from The UK National balance sheet estimates: 2021 and series YBHA and HABN (National Statistics, 2022); **265 European Commission:** Based on data in AMECO Database (European Commission, DGECFIN, May 2022); **268 Office for National Statistics:** UK National Accounts, The Blue Book: 2021 (ONS, October 2021); **269 Office for National Statistics:** UK National Accounts, The Blue Book: 2017, ONS, www.ons.gov.uk/economy/grossdomesticproductgdp/compendium/unitedkingdomnationalaccountsthebluebook/2020; **269 Office for National Statistics:** UK National Accounts, The Blue Book: 2017, ONS, www.ons.gov.uk/economy/grossdomesticproductgdp/compendium/unitedkingdomnationalaccountsthebluebook/2020; **270 International Monetary Fund:** World Economic Outlook Database (IMF, April 2022); **271 International Monetary Fund:** World Economic Outlook Database (IMF, April 2022); **284 Bank of England:** Based on data from Bank of England available at http://www.bankofengland.co.uk/publications/quarterlybulletin/threecenturiesofdata.xls; **284 Office for National Statistics:** Based on data from Quarterly National Accounts series ABMI and IHYP (National Statistics), www.ons.gov.uk/economy/grossdomesticproductgdp/; **286 European Commission:** Based on data from AMECO database (European Commission, DGECFIN); **288 Office for National Statistics:** Based on data in OECDStat (OECD, 2022); **289 Office for National Statistics:** Based on data in OECDStat (OECD, 2022), rebased by the authors; **289 Office for National Statistics:** Based on data in OECDStat (OECD, 2022); **290 Office for National Statistics:** Based on data from Capital Stocks, Consumption of Fixed Capital in the UK dataset (National Statistics), https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/bulletins/capitalstocksconsumptionoffixedcapital/2021; **291 European Commission:** Based on data from AMECO database (European Commission); **292 Office for National Statistics:** Based on series MLR3 and NPQT (National Statistics); **299 Office for National Statistics:** Based on data in Human Capital Estimates in the UK 2004 to 2020 (Office for National Statistics, April 2022); **306 Bank of England:** Based on data in Bankstats (Bank of England), Table B1.4, Data published 1 July 2022; **308 Bank of England:** (i) Data showing liabilities of banks and building societies based on series LPMALOA and RPMTBJF (up to the end of 2009) and RPMB3UQ (from 2010) from Statistical Interactive Database, Bank of England (data published 1 October 2018, not seasonally adjusted).; **308 Office for National Statistics:** (ii) GDP based on series YBHA, Office for National Statistics (GDP figures are the sum of the latest four quarters).; **314 Bank of England:** Based on data from Statistical Interactive Database, Bank of England, series LPMVUJD (up to 2010); **314 Office for National Statistics:** Based on data from LPMB8GO (data published 30 August 2018) and series YBHA, National Statistics.; **321 Bank of England:** Based on data from Statistical Interactive Database, series IUMABEDR and IUMASOIA (Bank of England) (data published I July 2022); **324 Bank of England:** Based on series LPQAUYN (M4), LPMAVAE (M0) until 2006, LPMBL22 (reserves) and LPMAVAB (notes and coin) from 2006, Statistical Interactive Database, Bank of England (data published 1 July 2022, seasonally adjusted except for reserves).; **325 Bank of England:** Statistical Interactive Database (Bank of England), Series LPMVQJW (data published 1 July 2022, seasonally adjusted); **327 Bank of England:** Based on data in Bankstats (Bank of England), Table A3.2 (Data published 29 March 2022); **328 Bank of England:** Based on Statistical Interactive Database, Bank of England, series LPQVWNV and LPQVWNQ; **328 Office for National Statistics:** Based on series YBHA, Office for National Statistics; **334 Bank of England:** Based on data from Statistical Interactive Database, series IUMABEDR, IUMASOIA, IUMZID2, IUMALNZC, IUMTLMV, Bank of England (1 July 2022); **343 European Commission:** Based on data in AMECO Database (European Commission) to 1980 and World Economic Outlook (IMF, April 2022); **343 European Commission:** AMECO database (European Commission, Economic and Financial Affairs, May 2022); **345 European Commission:** Based on data from Employment and Unemployment (LFS) (Eurostat, European Commission, 2022); **347 Office for National Statistics:** Based on series YBWF, YBWG and YBWH (Office for National Statistics); **351 European Commission:** Based on data in EMP16: Underemployment and overemployment (ONS, May 2022); **352 European Commission:** Based on data in data in AMECO Database (European Commission) (to 1981) and World Economic Outlook (IMF, April 2022); **353 Office for National Statistics:** Based on Time Series Data, series L55O and K8IA (ONS); **354 Office for National Statistics:** Based on Time Series Data, series D7G7, L55O, L5LQ, and K8IA (ONS); **360 European Commission:** Based on data in AMECO Database, (European Commission, DGECFIN, May 2022) and various forecasts; **362 Office for National Statistics:** Based on data from the Office for National Statistics; **362 International Monetary Fund:** Based on data from World Economic Outlook Database (IMF); **376 European Commission:** Based on data from AMECO database, Tables 16.3 and 18.1 (European Commission, DG ECFIN); **377 The National Archives:** Public Finances Databank, Office for Budget Responsibility (April 2022); **379 The National Archives:** Public Finances Databank, Office for Budget Responsibility (April 2022);

380 European Commission: Based on data from Fiscal Monitor, IMF eLibrary-Data, (IMF, April 2022); **382 The National Archives:** Based on data from Fiscal Monitor, IMF eLibrary-Data, (IMF, April 2022); **386 European Commission:** Based on data in World Economic Outlook Database, (IMF, April 2022); **403 International Monetary Fund:** Based on data from World Economic Outlook Database (IMF); **403 European Commission:** Based on AMECO database (European Commission); **406 Organisation for Economic Co-operation and Development:** Gross Domestic Spending on R&D (OECD, 2022); **415 The World Bank Group:** Series NY.GDP.DEFL.KD.ZG, World Bank, https://data.worldbank.org/indicator/NY.GDP.DEFL.KD.ZG; **416 International Monetary Fund:** Based on data from World Economic Outlook database (IMF, April 2022); **419 World Trade Organization:** Based on data from UNCTADStat (UNCTAD, 2022); **419 World Trade Organization:** Based on data in WTO Stats (WTO, 2022); **420 World Trade Organization:** Based on data in WTO Stats (WTO, 2022); **420 World Trade Organization:** Trade Profiles, Statistics Database (WTO, 2022); **423 European Commission:** Based on data in AMECO Database, (European Commission, DGECFIN, May 2022); **430 Pascal Lamy:** Global policy without democracy', speech by Pascal Lamy, given in 2001 when he was the EU Trade Commissioner. He later became head of the WTO in 2005; **440 The National Archives:** Adapted from EU Referendum: HM Treasury analysis key facts, HM Treasury (18 April 2016) (available at https://www.gov.uk/government/news/eu-referendum-treasury-analysiskey-facts); **441 Office for National Statistics:** Based on data from ONS Time Series Data, series IKBK, IKBL, ABMI, www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ikbk/ukea; **441 Organisation for Economic Co-operation and Development:** The economic consequences of Brexit: a taxing decision, OECD (25 April 2016) (available at http://www.oecd.org/economy/the-economic-consequences-of-brexit-ataxing-decision.htm).; **444 The World Bank Group:** Based on data from World Bank Commodity Price Data (The Pink Sheet), (Commodity Markets, World Bank); **448 The World Bank Group:** Based on data from Databank, World Bank (2022); **448 World Scientific:** Paul Krugman, 'Increasing Returns in a Comparative Advantage World', in Robert M.Stern, Comparative Advantage, Growth, and the Gains from Trade and Globalization (World Scientific, 2011) Chapter 7, p. 45.; **449 The Washington Post:** Jia Lynn Yang, 'China's manufacturing sector must reinvent itself, if it's to survive', The Washington Post (23 November 2012).; **454 Office for National Statistics:** Based on data from Balance of Payments time series and series YBHA (Office for National Statistics), www.ons.gov.uk/economy/nationalaccounts/balanceofpayments/datasets/balanceofpayments; **455 European Commission:** Based on data in AMECO Database (European Commission) (to 1980) and World Economic Outlook (IMF, April 2022); **457 Office for National Statistics:** Based on data from Balance of Payments time series and series YBHA (ONS); **458 Office for National Statistics:** Based on data in Statistical Interactive Database (Bank of England); **459 Bank for International Settlements:** Based on data from Effective exchange rate indices (Bank for International Settlements), Based on data from Effective exchange rate indices, Bank for International Settlements, www.bis.org/statistics/eer.htm; **468 Office for National Statistics:** Based on data from ons, series THAP and AUSS; **468 European Central Bank:** Based on the Data from Federal Reserve Bank and European Central Bank; **476 European Commission:** Based on data from AMECO database (European Commission, DGECFIN); **476 European Commission:** Based on data in AMECO Database (European Commission, DGECFIN); **477 European Commission:** Based on data in AMECO Database (European Commission, DGECFIN, May 2022); **478 European Commission:** Based on data in AMECO database (European Commission, DGECFIN); **479 The World Bank Group:** Based on data from Databank (World Bank), https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS; **480 The World Bank Group:** Based on data from Databank (World Bank), series DT.TDS.DECT.GN.ZS; **483 Organisation for Economic Co-operation and Development:** Based on data from Net ODA, OECD Data (OECD, 2022).

Photo credits

FM5 Dr Dean Garratt: Courtesy of Dr Dean Garratt; **FM5, 1, 2, 29, 30, 48, 71, 93, 121, 162, 193, 231, 232, 273, 301, 341, 374, 413, 414, 453 John Sloman:** Courtesy of John Sloman.

Part A

Introduction

Chapter 1

Economic issues

We start by looking at three of the biggest issues of our time – the COVID-19 pandemic, the subsequent rise in global inflation and global warming. These have had profound effects on societies around the world and, in the case of global warming, will do for decades to come. They are forcing us and our governments to make choices. Studying these choices is central to economics. Economists can analyse them and present us with policy alternatives. They can help us come to the best decisions in the light of the information presented by scientists.

Economics contains some core ideas. These ideas are simple but can be applied to a wide range of economic problems. We start examining these ideas in this chapter. We begin on the journey to help you to 'think like an economist' – a journey that we hope you will find fascinating and will give you a sound foundation for many possible future careers.

In this chapter, we will attempt to answer the question, 'What is economics about?', and give you greater insight into the subject you are studying. We will see how the subject is divided up and distinguish between the two major branches of economics: microeconomics and macroeconomics.

We will also look at the ways in which different types of economy operate, from the centrally planned economies of the former communist countries to the more free-market economies of most of the world today. We will ask just how 'markets' work.

After studying this chapter, you should be able to answer the following questions:

- What is economics about?
- What is the central economic problem faced by all individuals and societies?
- How can people set about making the best of their limited resources?
- What is meant by 'opportunity cost'? How is it relevant when people make economic choices?
- What is the difference between microeconomics and macroeconomics?
- What are the potential social implications of economic choices?
- How can you represent simply economic relationships in a graph?
- Why is the distinction between nominal and real figures important?
- How do different economic systems tackle the problem of scarcity?

At this point it's worth drawing your attention to the Economics News site that accompanies this text. You can access it directly at http://pearsonblog.campaignserver.co.uk/ or from MyLab Economics' home page, or simply Google 'Sloman Economics news site'. The site shows how items in the news are related to the economic issues you will be studying in this text. There are links to newspaper articles, to videos, to data sources and to reports. There are questions for you to consider and a powerful search feature that lets you browse earlier articles by chapter of the book, month and keywords.

1.1 GLOBAL ECONOMIC ISSUES

Problems to analyse and tackle

Economists study the various choices and decisions that affect our lives and livelihoods. These may be big decisions made by governments or the day-to-day decisions that we each make. To give you a flavour of economic decisions, we look at three of the most important issues that have faced societies and individuals in recent years.

COVID-19 and the global health emergency

The COVID-19 pandemic dominated our lives during 2020 and 2021 and beyond. People and governments struggled to cope with illness and death, and the damage to lives and livelihoods. Everyone was faced with choices, and these affected behaviour. Most of these had an economic dimension.

Individual choices

People had to decide whether to follow the rules and advice about behaviour: e.g. whether to wear a face-mask, or socially distance or follow lockdown rules. Economics studies people's behaviour – and how it impacts on economic decisions and the economy. We look at such behaviour in Chapters 2 to 4. For example, early on in the pandemic, many people stockpiled various items, such as hand sanitiser, toilet rolls and dried foods. This caused many shops to run out, which only further encouraged panic buying. Some shops responded by raising prices to increase their profit margins.

The lockdowns affects firms' profits. Some sectors were particularly hard hit, such as hospitality, leisure and tourism. Profits would have become losses if the government had not provided substantial support, which was still not enough to prevent many firms going out of business.

And the pandemic hastened the move to online sales and away from the High Street. Across the UK, some 17 500 chain-store outlets were permanently closed in 2020. In contrast, sales of online retailers such as Boohoo and Asos boomed.

We examine costs, revenues and profits in Chapters 5 and 6. We see how some firms are better protected against market forces than others, especially if they have a large market share and resulting market power.

As far as employees were concerned, some were easily able to work from home with a separate room to work in and a good Internet connection. They also saved money on commuting costs. Others with child-care responsibilities and shared working spaces and/or devices struggled to work efficiently from home. Some found their incomes constant or even rising; others saw a fall or had to rely on furlough money from the government.

Then vaccines began to be rolled out. Most people embraced getting jabbed to protect them and their loved ones. Others were suspicious for various reasons. But here was a classic problem in economics: what we do for ourselves often has spillover effects on others. If we are not protected, we are more likely to catch the disease and pass it on to others, even if we only get infected mildly or are largely asymptomatic. Many actions we take affect others – either beneficially or adversely.

Government choices

The pandemic did not just affect individuals and firms; it had major effects on whole economies. With many firms being forced to shut down, even if only temporarily, and some sectors, such as public transport, facing a collapse in demand, economies around the world went into recession – economic growth was negative and unemployment rose.

Governments in many countries responded by supporting individuals through various furlough schemes. Support was also given to businesses and to the self-employed. This prevented unemployment from rising much further.

Other longer-term measures for recovery included large-scale spending on physical infrastructure, such as public transport, roads, green energy and broadband, and on public services, such as health and education.

We look at issues such as growth and unemployment in Part C of the book.

But the massive support came at a cost. Government spending on support schemes plus a decline in tax revenues meant that government borrowing soared. Governments had to finance the borrowing through paying interest from taxes (or even more borrowing). So they were faced with the hard choice about when

to start raising taxes or cutting government spending to reduce the level of borrowing. The general approach was to spend now and pay later – an easy choice at the time, but a difficult one later, especially for governments facing re-election. Policy choices such as these are examined in Chapter 13.

Pause for thought

Give some other examples of choices that governments had to make during the pandemic. To what extent were they economic choices?

Global inflation

It was not just governments that were trying to keep their economies going during the pandemic. Central banks, such as the Federal Reserve in the USA, the European Central Bank for the eurozone and the Bank of England for the UK, were also playing their part. The general approach was to create more electronic money, through a process of 'quantitative easing'. If there was more money circulating through the banking system, people would borrow and spend more, helping to boost businesses.

But when you turn on the 'money tap' like this, you have to choose how much money to create and when to turn the tap off. Too little money and the recession may persist; too much money and prices may be pushed up by soaring spending. This 'inflation', as it is called, creates other problems for the economy, and central banks are keen not to let prices rise by more than 2 per cent per annum. The role of money in the economy is examined in Chapters 11–13.

Inflation really took off in 2022. The extra money created during the pandemic helped to fund the bounce back from recession. Spending rose rapidly. But supply could not keep up with the demand. Supply chains had been disrupted during the recession: there was a shortage of transportation (both land transport and shipping), a shortage of various components (such as computer chips) and a shortage of various types of skilled labour. Firms responded by raising prices. This pushed up the general level of prices: inflation took off.

Inflation was compounded by the Russian invasion of Ukraine. Supplies of oil and gas were disrupted; energy prices soared. Ukraine is a major exporter of wheat, sunflower oil and other crops. The blockade of Ukrainian exports led to a rapid rise in their price; food prices also soared. This further pushed up the overall inflation rate, which approached an annual rate of 10 per cent in many rich countries and higher rates than this in many poorer countries.

Central banks responded by raising interest rates – their normal response to rising inflation. The aim is to dampen spending by making borrowing more expensive and to encourage saving. But this risked driving countries into recession.

Inequality

A major issue in economics is inequality and poverty. Over the years this has increased in most countries as the rich have got richer, while many on low incomes have seen their real incomes (i.e. incomes after taking inflation into account) stagnate or even fall.

The inflation of 2022 compounded the problem. The poor consume a larger proportion of basic foodstuffs and energy than do the rich. Many poor people were driven into a state or food and fuel poverty and were not able to afford both to eat enough and to heat their homes adequately. We consider inequality in Chapter 7.

The environment and the global climate emergency

The world is faced with a climate emergency as the planet warms and as floods, droughts, fires, hurricanes and crop failures increase. Economists can look at policy options and their implications.

A key policy concerns pricing. If renewable energy were cheaper and fossil fuels were more expensive, then people would be more willing to switch to low-carbon consumption. Indeed, pricing is a central issue in economics. We look at pricing in Chapters 2, 3 and 5.

But how can prices be altered? They can be reduced by government subsidies and raised by taxes. We look at green taxes and subsidies in Chapter 8 and especially Section 8.9.

Another method for limiting emissions is for governments to cap the amount that firms are allowed to emit. Permits to emit CO_2 are allocated or auctioned to businesses. These can then be traded in markets. Low emitters will not have to pay so much, thereby giving them a cost advantage over high-emitting companies, which will require more permits and hence have to pay more. Economists have played a key role in developing emissions trading in markets such as the EU Emissions Trading Scheme (EU ETS).

International action

We live in an interdependent world. Actions in one part of the globe affect lives in others. A good example is cutting down rainforests for mining, ranching or growing monocrops, such as palm oil.

Actions by the global community can help but very often there are international games being played, with countries unwilling to commit to carbon-reducing measures unless they can be convinced that other countries are playing their part too. Economists study these types of 'games'. Indeed there is a major branch of economics called 'game theory', which looks at effective ways of incentivising people, firms and governments to behave in co-operative ways. We look at game theory in Section 6.6.

Pause for thought

For what reasons may governments want other governments to stick to tough climate or emissions targets and yet be not willing to do so themselves?

Trade between nations can make everyone better off. But this only works if certain conditions hold, including recognition of the environmental impact of trade. Economists study these conditions and can advise governments on trade policy. This and other international issues are the subject of the final two chapters of the book.

All these economic issues stem from a core set of problems. It is to this core that we now turn.

1.2 THE CORE OF ECONOMICS

What is economics all about?

Section 1.1 illustrates that economics involves an analysis of decision making by individuals, businesses, governments and countries. These are decisions concerned with the following:

- The **production** of goods and services: how much an economy produces, both in total and of individual items; how much each firm or person produces; what techniques of production are used; how many people are employed.
- The **consumption** of goods and services: how much people spend (and how much they save); how much people buy of particular items; what individuals choose to buy; how consumption is affected by prices, advertising, fashion, people's incomes and other factors.

But we still have not quite got to the bottom of what economics is about. What is the crucial ingredient that makes a problem an *economic* one? The answer is that there is one central problem faced by all individuals and all countries, no matter how rich. From this one problem stem all the other economic problems we shall be looking at throughout this text.

This central economic problem is *scarcity*. For an economist, scarcity has a very specific definition. Let us examine that definition.

The problem of scarcity

Ask people if they would like more money, and the vast majority would answer 'yes'. They want more money so that they can buy more goods and services; and this applies not only to poor people but also to most wealthy people too. The point is that human wants are virtually unlimited.

Yet the means of fulfilling human wants are limited. At any one time the world can produce only a limited amount of goods and services. This is because the world has only a limited amount of resources. These resources, or **factors of production** as they are often called, are of three broad types.

- Human resources: **labour**. The labour force is limited in number and in skills.
- Natural resources: **land and raw materials**. The world's land area is limited, as are its raw materials
- Manufactured resources: **capital**. Capital consists of all those inputs that have themselves been produced in the first place. The world has a limited stock of capital: a limited supply of factories, machines, transportation and other equipment. The productivity of capital is limited by the state of technology.

So here is the reason for scarcity: human wants are virtually unlimited, whereas the resources available to satisfy these wants are limited. We can thus define **scarcity** as follows:

KEY IDEA 1

Scarcity **is the excess of human wants over what can actually be produced. Because of scarcity, various choices have to be made between alternatives.**

Of course, we do not all face the problem of scarcity to the same degree. A poor person unable to afford enough to eat or a decent place to live will hardly see it as a 'problem' that a rich person cannot

Definitions

Production The transformation of inputs into outputs by firms in order to earn profit (or meet some other objective).

Consumption The act of using goods and services to satisfy wants. This will normally involve purchasing the goods and services.

Factors of production (or resources) The inputs into the production of goods and services: labour, land and raw materials, and capital.

Labour All forms of human input, both physical and mental, into current production.

Land (and raw materials) Inputs into production that are provided by nature: e.g. unimproved land and mineral deposits in the ground.

Capital All inputs into production that have themselves been produced: e.g. factories, machines and tools.

Scarcity The excess of human wants over what can actually be produced to fulfil these wants.

afford a second Ferrari. But economists do not claim that we all face an *equal* problem of scarcity. In fact, this is one of the major issues economists study: how resources and products are *distributed*, whether between different individuals, different regions of a country or different countries of the world.

Pause for thought

If we would all like more money, why doesn't the government simply print a lot more?

But given that people, both rich and poor, want more than they can have, this makes them *behave* in certain ways. Economics studies that behaviour. It studies people at work, producing the goods that people want. It studies people as consumers buying the goods they themselves want. It studies governments influencing the level and pattern of production and consumption. In short, it studies anything to do with the process of satisfying human wants.

Demand and supply

We said that economics is concerned with consumption and production. Another way of looking at this is in terms of *demand* and *supply*. In fact, demand and supply and the relationship between them lie at the very centre of economics. But what do we mean by the terms, and what is their relationship with the problem of scarcity?

Demand is related to wants. If goods and services were free, people would simply demand whatever they wanted. Such wants are virtually boundless, perhaps limited only by people's imagination. *Supply*, on the other hand, is limited. It is related to resources. The amount firms can supply depends on the resources and technology available.

Given the problem of scarcity, given that human wants exceed what can actually be produced, *potential* demands will exceed *potential* supplies. Society therefore has to find some way of dealing with this problem. Somehow it has to try to match demand and supply. This applies at the level of the economy overall: total spending in the economy must balance total production. It also applies at the level of individual goods and services. The demand and supply of cabbages must balance, and so must the demand and supply of cars, houses, tablets and holidays.

KI 1 p 5

But if potential demand exceeds potential supply, how are *actual* demand and supply to be made equal? Either demand has to be curtailed, or supply has to be increased, or a combination of the two. Economics studies this process. It studies how demand adjusts to available supplies, and how supply adjusts to consumer demands.

Recap

1. The central economic problem is that of scarcity.
2. Given that there is a limited supply of factors of production (labour, land and capital), it is impossible to provide everybody with everything they want.
3. Potential demands exceed potential supplies.

1.3 DIVIDING UP THE SUBJECT

What's meant by 'microeconomics' and 'macroeconomics'?

Economics is traditionally divided into two main branches – *microeconomics and macroeconomics*, where 'micro' means small and 'macro' means big.

Microeconomics is concerned with the individual parts of the economy. It is concerned with the demand and supply of *particular* goods and services and resources such as cars, butter, clothes, haircuts, plumbers, accountants, blast furnaces, computers and oil.

Macroeconomics is concerned with the economy as a whole. It is thus concerned with **aggregate demand**

Definition

Microeconomics The branch of economics that studies individual units: e.g. households, firms and industries. It studies the interrelationships between these units in determining the pattern of production and distribution of goods and services.

and **aggregate supply.** By 'aggregate demand' we mean the total amount of spending in the economy, whether by consumers, by customers outside the country for our exports, by the government, or by firms when they buy capital equipment or stock up on raw materials. By 'aggregate supply' we mean the total national output of goods and services.

Microeconomics

Microeconomics and choice

KI 1 p 5

Because resources are scarce, *choices* have to be made. There are three main categories of choice that must be made in any society.

- *What* goods and services are going to be produced and in what quantities? How many cars, how much wheat, how much insurance, how many rock concerts, etc. will be produced?
- *How* are things going to be produced? What resources are going to be used and in what quantities? What techniques of production are going to be adopted? Will cars be produced by robots or by assembly-line workers? Will electricity be produced from coal, oil, gas, nuclear fission, renewable resources or a mixture of these?
- *For whom* are things going to be produced? In other words, how will the nation's income be distributed? After all, the higher your income, the more you can consume of the nation's output. What will be the wages of shop workers, professional footballers, cleaners and accountants? How much will chief executives of large companies receive? How much will pensioners receive? How much of the nation's income will go to shareholders or landowners?

All societies have to make these choices, whether they be made by individuals, by groups or by the government. These choices are *micro*economic choices, since they are concerned not with the *total* amount of national output, but with the *individual* goods and services that make it up: what they are, how they are made, and who gets the incomes to buy them.

Choice and opportunity cost

Choice involves sacrifice. The more food you choose to buy, the less money you will have to spend on other goods. The more food a nation produces, the less resources there will be for producing other goods. In other words, the production or consumption of one thing involves the sacrifice of alternatives. This sacrifice of alternatives in the production (or consumption) of a good is known as its **opportunity cost**.

If the workers on a farm can produce either 1000 tonnes of wheat or 2000 tonnes of barley, then the opportunity cost of producing 1 tonne of wheat is the 2 tonnes of barley forgone. The opportunity cost of buying a textbook is the new pair of jeans you also wanted that you have had to go without. The opportunity cost of working overtime is the leisure you have sacrificed.

Opportunity cost as the basis for choice is a key idea. But it is more than that. It is also the first of our 'threshold concepts' (click on the Threshold Concepts link in MyLab Economics or on the student website for a detailed explanation of each one). There are 15 of these threshold concepts, which we shall be exploring throughout the text. Each of them keeps recurring in a variety of different contexts.

Once you have grasped these concepts and seen their significance, they will affect the way that you understand and analyse economic problems. They help you to 'think like an economist'.

KEY IDEA 2 TC 1

The *opportunity cost* of something is what you give up to get it/do it.

Rational choices

When trying to understand behaviour economists typically start by assuming 'rational decision making'. Consequently, they often refer to ***rational choices***. This simply means the weighing-up of the *costs* and *benefits* of any activity, whether it be firms choosing what and how much to produce, workers choosing whether to take a particular job or to work extra hours, or consumers choosing what to buy.

Imagine you are doing your shopping in a supermarket and you want to buy a bottle of wine. Do you

Definitions

Macroeconomics The branch of economics that studies economic aggregates (grand totals): e.g. the overall level of prices, output and employment in the economy.

Aggregate demand The total level of spending in the economy.

Aggregate supply The total amount of output in the economy.

Opportunity cost The cost of any activity measured in terms of the best alternative forgone.

Rational choices Choices that involve weighing up the benefit of any activity against its opportunity cost.

BOX 1.1 **THE OPPORTUNITY COSTS OF STUDYING** CASE STUDIES & APPLICATIONS

What are you sacrificing?

You may not have realised it, but you probably consider opportunity costs many times a day. The reason is that we are constantly making choices: what to buy, what to eat, what to wear, whether to go out, how much to study and so on. Each time we make such a choice, we are in effect rejecting some alternative. This alternative forgone is the opportunity cost of the action we chose.

TC 1 p 7

Sometimes the opportunity costs of our actions are the direct monetary costs we incur. Sometimes it is more complicated.

Take the opportunity costs of your choices as a student.

Buying a textbook costing £49.95

This choice involves a direct money payment. What you have to consider is the alternatives you could have bought with the £49.95. You then have to weigh up the benefit from the best alternative against the benefit of the textbook.

1. *What might prevent you from making the best decision?*

Coming to classes

Even though students pay fees for their degrees in many countries, there is no extra (marginal) monetary cost in coming to classes once the fees have been paid. You will not get a refund by missing classes. The fees, once you've paid them, are what we call a 'sunk cost'.

So are the opportunity costs zero? No: by coming to classes you are *not* working in the library; you are *not* having an extra hour in bed; you are *not* undertaking paid work during that time, and so on. If you are making a rational decision to come to classes, then you will consider such possible alternatives.

2. *If there are several other things you could have done, is the opportunity cost the sum of all of them?*
3. *What factors would make the opportunity cost of attending a class relatively high?*

Revising for an economics exam

Again, the opportunity cost is the best alternative to which you could have put your time. This might be revising for some other exam. You will probably want to divide your time sensibly between your subjects. A *sensible* decision is not to revise economics on any given occasion if you will gain a greater benefit from revising another subject. In such a case the (marginal) opportunity cost of revising economics exceeds the (marginal) benefit.

Choosing to study at university or college

What are the opportunity costs of being a student in higher education? At first it might seem that the costs would include the following:

- Tuition fees.
- Books, stationery, etc.
- Accommodation, food, entertainment, travel and other living expenses.

But adding these up does *not* give the *opportunity* cost. The opportunity cost is the *sacrifice* entailed by going to university or college *rather* than doing something else. Let us assume that the alternative is to take a job that has been offered. The correct list of opportunity costs of higher education would include:

- Books, stationery, etc.
- Additional accommodation and travel expenses over what would have been incurred by taking the job (this figure could be negative).
- Wages that would have been earned in the job, less any income received as a student.
- Tuition fees paid by the student.

4. *Why is the cost of food not included?*
5. *What impact would it have on the calculation of opportunity costs if you really disliked the nature of the work in the best alternative job?*
6. *Is the opportunity cost to the individual of attending higher education different from the opportunity costs to society as a whole? Do the benefits of higher education for society differ from those for the individual?*

Estimate your own cost of studying for a degree (or other qualification). For what reasons might you find it difficult to make such a calculation?

spend a lot of money and buy a top-quality wine, or do you buy a cheaper one instead? To make a rational (i.e. sensible) decision, you will need to weigh up the costs and benefits of each alternative. The top-quality wine may give you a lot of enjoyment, but it has a high opportunity cost: because it is expensive, you will need to sacrifice quite a lot of consumption of other goods if you decide to buy it. If you buy the cheap bottle, however, although you will not enjoy it so much, you will have more money left over to buy other things: it has a lower opportunity cost.

TC 1 p 7

Thus rational decision making, as far as consumers are concerned, involves choosing those items that give you the best value for money: i.e. the *greatest benefit relative to cost*.

The same principles apply to firms when deciding what to produce. For example, should a car manufacturer open up another production line? A rational decision will again involve weighing up the benefits and costs. The benefits are the revenues that the firm will earn from selling the extra cars. The costs will include the extra labour costs, raw material costs, costs of component parts, etc. It will be profitable to

open up the new production line only if the revenues earned exceed the costs entailed: in other words, if it increases profit.

Marginal costs and benefits

In economics we argue that rational choices involve weighing up **marginal costs** and **marginal benefits**. These are the costs and benefits of doing a little bit more or a little bit less of a specific activity. They can be contrasted with the *total* costs and benefits of the activity.

TC 2 p 9

Take a familiar example. What time will you set the alarm clock to go off tomorrow morning? Let us say that you have to leave home at 8.30 a.m. Perhaps you will set the alarm for 7 a.m. That will give you plenty of time to get up and get ready, but it will mean a relatively short night's sleep. Perhaps then you will decide to set it for 7.30 a.m. or even 8 a.m. That will give you a longer night's sleep, but much more of a rush in the morning to get ready.

So how do you make a rational decision about when the alarm should go off? What you have to do is to weigh up the costs and benefits of *additional* sleep. Each extra minute in bed gives you more sleep (the marginal benefit) but gives you more of a rush when you get up (the marginal cost). The decision is therefore based on the costs and benefits of *extra* sleep, not on the *total* costs and benefits of a whole night's sleep.

This same principle applies to rational decisions made by consumers, workers and firms. For example, the car firm we were considering just now will weigh up the marginal costs and benefits of producing cars: in other words, it will compare the costs and revenue of producing *additional* cars. If additional cars add more to the firm's revenue than to its costs, it will be profitable to produce them.

KEY IDEA 3

TC 2

Rational decision making **involves weighing up the marginal benefit and marginal cost of any activity. If the marginal benefit exceeds the marginal cost, it is rational to do the activity (or to do more of it). If the marginal cost exceeds the marginal benefit, it is rational not to do it (or to do less of it).**

Decision making based on marginal costs and benefits is the second of our threshold concepts, explored on the book's website.

Definitions

Marginal costs The additional costs of doing a little bit more (or *1 unit* more if a unit can be measured) of an activity.

Marginal benefits The additional benefits of doing a little bit more (or *1 unit* more if a unit can be measured) of an activity.

Pause for thought

Imagine that, as a student, you are short of money and that you are offered employment working in the student union shop. You can choose the number of hours each week that you work. How would you make a 'rational' decision about the number of hours to work in any given week?

The social implications of choice

Microeconomics does not just study how choices are made. It also looks at their *consequences*. Under certain conditions the consequences may be an *efficient* use of the nation's resources: the economy is making the most of its scarce resources.

However, a whole series of possible problems can arise from the choices that people make, whether they are made by individuals, by firms or by the government. These problems include such things as inefficiency, waste, inequality and pollution.

Take the cases of inequality and pollution.

- *Inequality.* Even though the current levels of production and consumption might be efficient, they might be regarded as unfair. For the distribution of goods and services among different members of societies to be regarded as equitable it must be considered fair or just. The problem, of course, is that people have different notions of fairness. Equity is therefore described as a value judgement: notions of equity will depend on the values of individuals or society.
- *Pollution.* It might be profitable for a firm to tip toxic waste into a river. But what is profitable for the firm will not necessarily be 'profitable' for society. There may be serious environmental consequences of the firm's actions. The case of pollution illustrates how the effects of people's choices often spill over to other people.

Macroeconomics

Because things are scarce, societies are concerned that their resources should be used as *fully as possible*, and that over time their national output should *grow*.

KI 1 p 5

The achievement of growth and the full use of resources is not easy, however, as demonstrated by the periods of high unemployment and stagnation that have occurred from time to time throughout the world (for example, in the 1930s, the early 1980s, the early 1990s, the late 2000s and 2020–1 during the pandemic lockdowns). Furthermore, attempts by government to stimulate growth and employment can result in inflation and rising imports.

Even when societies do achieve growth, it can be short-lived. Economies are inherently unstable and

display what are known as **business cycles**: fluctuations in the economy's growth rates. The more that growth rates fluctuate, the more marked are the fluctuations in the path of output traced out by the economy over time. The inherent instability of economies is our next threshold concept.

KEY IDEA 4 TC 3

Economies suffer from inherent instability. **As a result, economic growth and other macroeconomic indicators tend to fluctuate. This is Threshold Concept 3. It is a threshold concept because it is vital to recognise the fundamental instability in market economies. Analysing the ups and downs of the 'business cycle' occupies many macroeconomists.**

Fluctuations in the overall level of economic activity are linked to changes in the total demand for firms' goods and services (aggregate demand) and/or in the amount that firms plan to supply at any given level of prices (aggregate supply). These changes affect the balance between aggregate demand and aggregate supply.

And it is not just economic growth rates: other macroeconomic problems too are closely related to the balance between aggregate demand and aggregate supply. If aggregate demand is *too low* relative to aggregate supply, recession and unemployment may well result.

- **Recession** is where the economy's output declines for two successive quarters or longer. In other words, during this period growth becomes negative. Hence, not all periods during which the economy contracts are termed 'recessions'. It is the duration and persistence of the contraction that distinguishes a recession.

  Recessions are associated with low levels of consumer spending. If people spend less, shops are likely to find themselves with unsold stocks. Then they will buy less from the manufacturers, who in turn will cut down on production.
- **Unemployment** is likely to result from cutbacks in production. If firms are producing less, they will need to employ fewer people.

If aggregate demand is *too high* relative to aggregate supply, inflation and balance of trade deficits are likely to result.

- **Inflation** refers to a general rise in the level of prices throughout the economy. If aggregate demand rises substantially, firms are likely to respond by raising their prices. If demand is high, they can probably still sell as much as before (if not more) even at the higher prices, and thus make higher profits. If firms in general put up their prices, inflation results. By comparing price levels over time, we can measure the **rate of inflation**. Typically, this is the *annual* rate of inflation: the percentage increase in prices over a 12-month period.
- **Balance of trade** deficits are the excess of imports over exports. If aggregate demand rises, people are likely to buy more imports. In other words, part of the extra expenditure will go on Japanese electrical goods, German cars, Chilean wine, and so on. Also, if the rate of inflation is high, home-produced goods will become uncompetitive with foreign goods. We are likely, therefore, to buy more imports, and people abroad are likely to buy fewer of our exports.

An important focus of government macroeconomic *policy*, therefore, is on the balance of aggregate demand and aggregate supply. It can be **demand-side policy**, which seeks to influence the level of spending in the economy. This in turn will affect the level of production, prices and employment. Or it can be **supply-side policy**. This is designed to influence the level of production directly: for example, by trying to create more incentives for firms to innovate.

Macroeconomic policies can also be used to foster future economic growth. Policies in this context are designed to increase future levels of production by raising the economy's productive capabilities. Therefore, supply-side policy is generally agreed to be particularly important, though if demand-side policy

Definitions

Business cycle The periodic fluctuations of a country's economic growth rates around the long-term trend.

Recession A period where national output falls for two or more successive quarters.

Unemployment The number of people who are actively looking for work but are currently without a job. (Note that there is much debate as to who should officially be counted as unemployed.)

Inflation A general rise in the level of prices throughout the economy

(Annual) Rate of inflation The percentage increase in the level of prices over a 12-month period.

Balance of trade Exports of goods and services minus imports of goods and services. If exports exceed imports, there is a 'balance of trade surplus' (a positive figure). If imports exceed exports, there is a 'balance of trade deficit' (a negative figure).

Demand-side policy Government policy designed to alter the level of aggregate demand, and thereby the level of output, employment and prices.

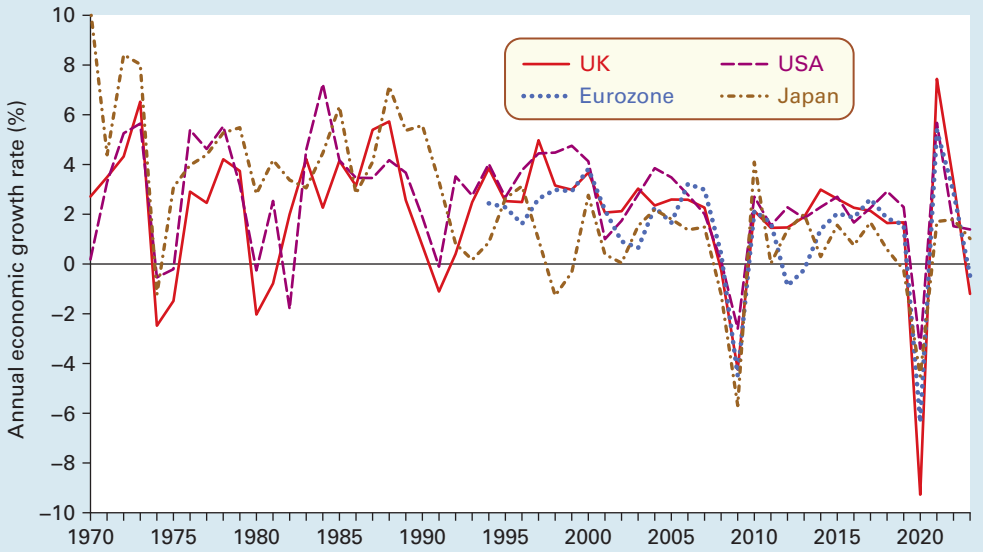
Supply-side policy Government policy that attempts to alter the level of aggregate supply directly.

BOX 1.2 **BUSINESS CYCLES** **EXPLORING ECONOMICS**

The inherent volatility of economies

Countries rarely experience stable economic growth. Instead they experience business cycles: periods of rapid economic growth followed by periods of low growth or even negative growth (falling output). The volatility of growth is readily seen when we plot annual percentage changes in the volume of output. The chart does this for four economies. In the case of the UK, the growth of output ranges from +6.5 per cent (1973) to −9.3 per cent in the pandemic (2020). In Japan, it ranges from +10.3 per cent (1970) to −5.7 per cent (2009).

Growth rates in selected industrial economies



Notes: 2022 and 2023 based on forecasts; eurozone figures are the weighted average of the countries using the euro in any given year; the euro was introduced in 1999; the eurozone figures before 1999 are the weighted average of the original members

Source: Based on data in *AMECO Database,* European Commission, DG Business, Economy, Euro (May 2022) and various forecasts for 2022 and 2023.

Explaining volatility

Some economists see the problem as rooted in fluctuations in *aggregate demand*: in other words, in the total demand for the economy's goods and services, whether by individuals or firms. Changes in the demands of individuals and firms may interact with each other, affecting the character of the business cycle.

For example, a rise in consumer spending could stimulate firms to invest to build up capacity to meet the extra demand. This, in turn, generates more employment, additional national income and so more consumption. A similar effect could occur if banks felt able to lend more in response to the growing economy, which would further stimulate the economy as these funds are spent.

Other economists see the problem as rooted in fluctuations in *aggregate supply*: in other words, in the total amount of goods and services firms plan to supply at a given level of prices. These economists argue that changes in aggregate supply occur if the price, availability or effectiveness of the inputs in firms' production processes are in some way affected.

One example could be technological changes that boost output and employment. Often such changes come in waves, which would contribute to the observed volatility of output. Another could be disruptions to supply chains caused by a pandemic or a war.

But whatever the cause, it is vital to recognise the fundamental instability of market economies. This is what makes the volatility in short-term growth rates a threshold concept.

1. *If people believe that the economy's output level is about to fall, how may their actions aggravate the problem?*
2. *Why will some people suffer more than others from a downturn in economic activity?*

Using the AMECO database, construct a line chart of the level of GDP at constant prices for the eurozone in € billions. Briefly summarise what the chart shows and the relationship between it and the eurozone series shown in the chart in this box.

were to reduce economic volatility, it too could contribute to future growth by encouraging investment in capital and labour.

How desirable is economic growth? Despite economic growth being a major objective of governments, it cannot guarantee rising levels of well-being. Economic growth can have an array of social implications.

An important illustration of this is the increasing emphasis placed on sustainable economic development. This recognises the importance of considering the impact of economic development on the environment. For economic development to be sustainable, it should not undermine the well-being of the planet's natural resources and ecosystems.

The significance of sustainable economic development has been formally recognised by the international community. In September 2015 the United Nations adopted the 2030 Agenda for Sustainable Development.[1] This saw a commitment to 17 Sustainable Development Goals (SDGs) and 169 targets. The SDGs aim to end poverty, tackle inequalities and ensure prosperity for all while protecting the planet. In other words, the SDGs formally recognise the importance of **sustainable economic growth**: growth that aims to satisfy human needs, but which sustains the planet's natural resources for future generations.

Another issue is who benefits from growth. It may be that faster economic growth results in greater *inequality*, with the rich seeing their incomes growing faster than those of the poor. Similarly, policies to redistribute incomes from rich to poor may result in slower economic growth. Governments may thus have to choose between faster economic growth and reductions in inequality.

TC 1 p 7

Recap

1. The subject of economics is usually divided into two main branches, macroeconomics and microeconomics.
2. Microeconomics deals with the activities of individual units within the economy: firms, industries, consumers, workers, etc. Because resources are scarce, people must make choices. Society has to choose by some means or other what goods and services to produce, how to produce them and for whom to produce them. Microeconomics studies these choices.
3. Rational choices involve weighing up the marginal benefits of each activity against its marginal opportunity costs. If the marginal benefits exceed the marginal costs, it is rational to choose to do more of that activity.
4. Macroeconomics deals with aggregates' such as the overall levels of unemployment, output, growth and prices in the economy.

1.4 MODELLING ECONOMIC RELATIONSHIPS

How can diagrams be used to illustrate economic issues?

Economics books and articles frequently contain diagrams. The reason is that diagrams are very useful for illustrating economic relationships. Ideas and arguments that might take a long time to explain in words can often be expressed clearly and simply in a diagram.

Such diagrams are an example of an **economic model**. Models are a simplification of reality in order to show clearly how things are related. For example, a model could show how the demand for a product is related to price. It would show how demand changed as price changed. We would expect that as the price of a product rose, so the quantity demanded would fall. If you glance forward to Figure 2.1 on page 32, you will see a diagram relating the demand for potatoes to the price of potatoes.

To examine such relationships, we have to assume that other things that might affect the outcome do not change. For example, if we were modelling how the demand for Cheddar cheese was affected by changes in its price, we would have to assume that the price of other products, including other types of cheese,

Definitions

Sustainable economic growth Economic growth which satisfies human needs, but which sustains the planet's natural resources for future generations.

Economic model The representation, either graphically, mathematically or in words, of the relationship between two or more variables. A model is a simplification of reality designed to explain just part of a complex process of economic relationships. It is thus based on various simplifying assumptions.

[1] See *Transforming our world: the 2030 Agenda for Sustainable Development* (United Nations, 21 October).

were held constant. This is known as the '***ceteris paribus***' assumption. *Ceteris paribus* is Latin for 'other things being equal'. If the price of other products *did* change, we would then have to re-specify our model.

Modelling in economics is the fourth of our threshold concepts.

KEY IDEA 5 TC 4

Modelling in economics involves specifying how one variable (the 'dependent variable') depends on one or more other variables ('independent variables'). It involves 'holding constant' all other variables that might influence the outcome (the *ceteris paribus* assumption). A model can be expressed in words, as a graph, or mathematically in terms of one or more equations. In this text we use mainly verbal descriptions and graphs.

Two of the most common types of diagram used in economics are graphs and flow diagrams. In the next two sections we will look at one example of each. These examples are chosen to illustrate the distinction between microeconomic and macroeconomic issues.

The production possibility curve

TC 4 p 13

We start by having a look at a **production possibility curve**. This diagram is a graph. Like many diagrams in economics it shows a simplified picture of reality – a picture stripped of all details that are unnecessary to illustrate the points being made. A production possibility curve is shown in Figure 1.1. The graph is based on the data shown in Table 1.1.

Assume that some imaginary nation devotes all its resources – land, labour and capital – to producing just two goods, food and clothing. Various possible combinations that could be produced over a given period of time (e.g. a year) are shown in the table. Thus the country, by devoting all its resources to producing food, could produce 8 million units of food but no clothing. Alternatively, by producing, say, 7 million units of food it could release enough resources – land, labour and capital – to produce 2.2 million units of clothing. At the other extreme, it could produce 7 million units of clothing, with no resources at all being used to produce food.

The information in the table can be transferred to a graph (Figure 1.1). We measure units of food on one axis (in this case the vertical axis) and units of clothing on the other. The curve shows all the combinations of the two goods that can be produced with all the nation's resources fully and efficiently employed. For example, production could take place at point *x*, with 6 million units of food and 4 million units of clothing being produced. Production cannot take place beyond the curve. For example, production is not possible at point *w*: the nation does not have enough resources to do this.

Note that there are two simplifying assumptions in this diagram. First, it is assumed that there are just two types of good that can be produced. It is necessary

TC 4 p 13

Figure 1.1 A production possibility curve

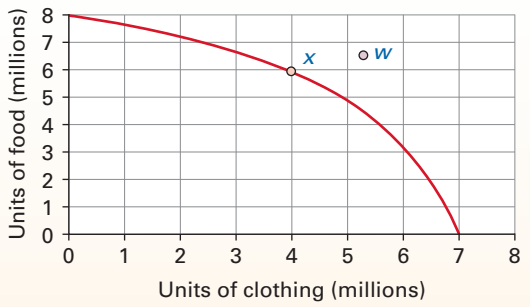


Table 1.1 Maximum possible combinations of food and clothing that can be produced in a given time period

Units of food (millions)	Units of clothing (millions)
8.0	0.0
7.0	2.2
6.0	4.0
5.0	5.0
4.0	5.6
3.0	6.0
2.0	6.4
1.0	6.7
0.0	7.0

Definitions

Ceteris paribus Latin for 'other things being equal'. This assumption has to be made when constructing and using models.

Production possibility curve A curve showing all the possible combinations of two goods that a country can produce within a specified time period with all its resources fully and efficiently employed.

to make this assumption since we only have two axes on our graph. The other assumption is that there is only one type of food and one type of clothing. This is implied by measuring their output in particular units (e.g. tonnes).

These two assumptions are obviously enormous simplifications when we consider the modern complex economies of the real world. But despite this, the diagram still allows important principles to be illustrated, and illustrated simply.

Microeconomics and the production possibility curve

A production possibility curve illustrates the microeconomic issues of *choice* and *opportunity cost*.

If the country chose to produce more clothing, it would have to sacrifice the production of some food. KI 1 p 5 This sacrifice of food is the opportunity cost of the extra clothing.

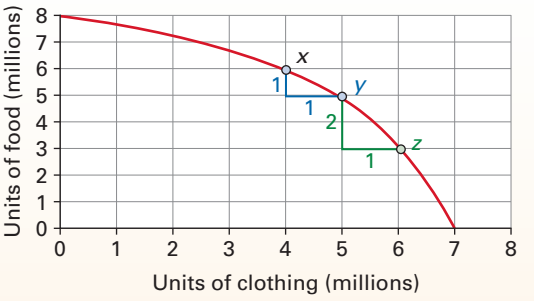
The fact that to produce more of one good involves producing less of the other is illustrated by the downward-sloping nature of the curve. For example, the country could move from point *x* to point *y* in Figure 1.2. In doing so it would be producing an extra 1 million units of clothing, but 1 million fewer units of food. Thus the opportunity cost of the 1 million extra units of clothing would be the 1 million units of food forgone.

It also illustrates the phenomenon of **increasing opportunity costs**. TC 1 p 5 By this we mean that as a country produces more of one good it has to sacrifice ever-*increasing* amounts of the other. The reason for this is that different factors of production have different properties. People have different skills. Land differs in different parts of the country. Raw materials differ one from another; and so on. Thus, as the nation concentrates more and more on the production of one good, it has to start using resources that are less and less suitable – resources that would have been better suited to producing other goods. In our example, then, the production of more and more clothing will involve a growing *marginal cost*: ever-increasing amounts of food have to be sacrificed for each additional unit of clothing produced.

Definition

Increasing opportunity costs When additional production of one good involves ever-increasing sacrifices of another.

Figure 1.2 Increasing opportunity costs

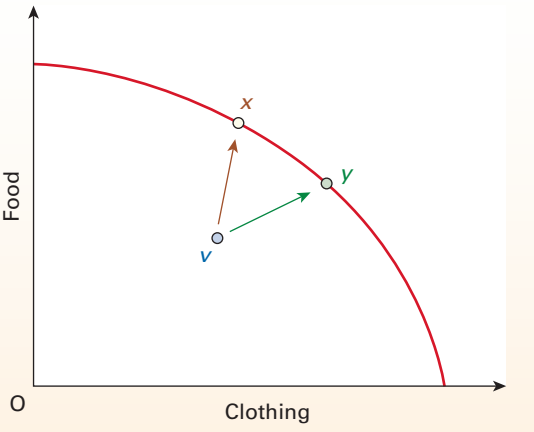


It is because opportunity costs increase that the production possibility curve is bowed outward rather than being a straight line. Thus in Figure 1.2 as production moves from point *x* to *y* to *z*, so the amount of food sacrificed rises for each additional unit of clothing produced. The opportunity cost of the fifth million units of clothing is 1 million units of food. The opportunity cost of the sixth million units of clothing is 2 million units of food.

Macroeconomics and the production possibility curve

There is no guarantee that resources will be fully employed, or that they will be used in the most efficient way possible. The nation may thus be producing at a point inside the curve: for example, point *v* in Figure 1.3.

Figure 1.3 Making a fuller use of resources



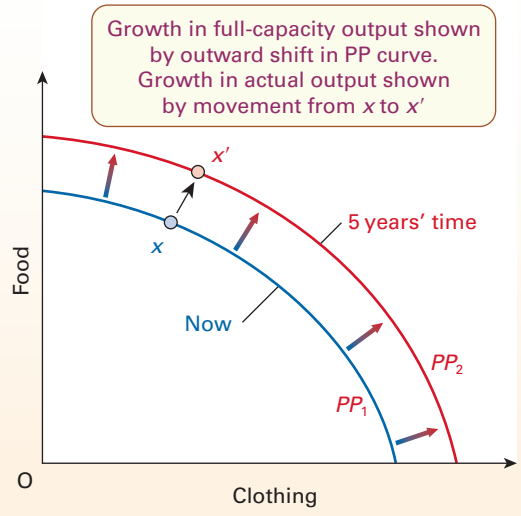
What we are saying here is that the economy is producing less of both goods than it could possibly produce, either because some resources are not being used (for example, workers may be unemployed), or because it is not using the most efficient methods of production possible, or a combination of the two. By using its resources to the full, however, the nation could move out on to the curve: to point x or y, for example. It could thus produce more clothing *and* more food.

Here we are not concerned with the combination of goods produced (a microeconomic issue), but with whether the total amount produced is as much as it could be (a macroeconomic issue).

As we move closer to an economy's full-capacity output and resources become increasingly scarce it is likely that inflationary pressures will increase. Therefore, it is unlikely that an economy could sustain working at full capacity beyond the short term. But, over time, the production possibilities of a nation are likely to increase. **Investment** in new plant and machinery will increase the stock of capital; new raw materials may be discovered; technological advances are likely to take place; through education and training, labour is likely to become more productive. This growth in full-capacity output is illustrated by an outward shift in the production possibility curve. This will then allow actual output to increase: for example, from point x to point x' in Figure 1.4.

TC4 p13

Figure 1.4 Growth in actual and full-capacity output



Pause for thought

Will economic growth necessarily involve a parallel outward shift in the production possibility curve? Explain.

The circular flow of goods and incomes

The process of satisfying human wants involves producers and consumers. The relationship between them is two-sided and can be represented in a flow diagram (see Figure 1.5).

The consumers of goods and services are labelled 'households'. Some members of households, of course, are also workers, and in some cases are the owners of other factors of production too, such as land. The producers of goods and services are labelled 'firms'.

Firms and households are in a twin 'demand and supply' relationship with each other.

First, in the top half of the diagram, households demand goods and services, and firms supply goods and services. In the process, exchange takes place. In a money economy (as opposed to a **barter economy**), firms exchange goods and services for money. In other words, money flows from households to firms in the form of consumer expenditure, while goods and services flow the other way – from firms to households.

This coming together of buyers and sellers is known as a **market** – whether it be a street market, a shop, the Internet, an auction, a mail-order system or whatever. Thus we talk about the market for apples, the market for oil, for cars, for houses, for televisions and so on.

Secondly, firms and households come together in the market for *factors of production*. This is illustrated in the bottom half of the diagram. This time, the demand and supply roles are reversed. Firms demand the use of factors of production owned by

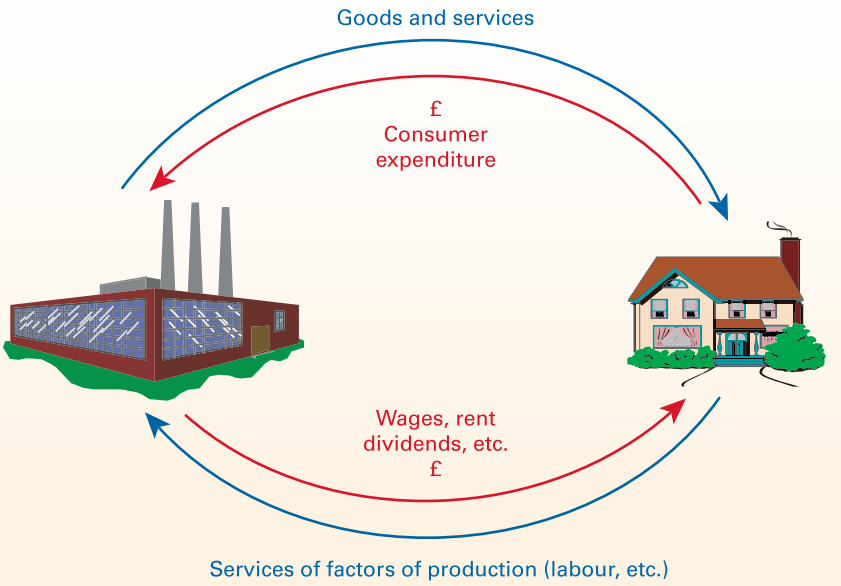
Definitions

Investment The production of items that are not for immediate consumption.

Barter economy An economy where people exchange goods and services directly with one another without any payment of money. Workers would be paid with bundles of goods.

Market The interaction between buyers and sellers.

Figure 1.5 Circular flow of goods and incomes



households – labour, land and capital. Households supply them. Thus the services of labour and other factors flow from households to firms, and in exchange firms pay households money – namely, wages, rent, dividends and interest. Just as we referred to particular goods markets, so we can also refer to particular factor markets – the market for bricklayers, for secretaries, for hairdressers, for land, and so on.

There is thus a circular flow of incomes. Households earn incomes from firms, and firms earn incomes from households. The money circulates. There is also a circular flow of goods and services, but in the opposite direction. Households supply factor services to firms, which then use them to supply goods and services to households.

The flow diagram in Figure 1.5, like the production possibility curve, can help us to distinguish between micro- and macroeconomics.

Microeconomics is concerned with the composition of the circular flow: *what* combinations of goods make up the goods flow; *how* the various factors of production are combined to produce these goods; *for whom* the wages, dividends, rent and interest are paid out.

Macroeconomics is concerned with the total size of the flow and what causes it to expand and contract.

Models and data

Economic models, such as the production possibility curve and the circular flow of income, provide us with simplified frameworks with which to analyse economic issues. Economic data (real or hypothetical) can be used to contextualise the issues addressed by our models, to quantity relationships identified by them, and to analyse further the implied outcomes. As you go through the text you will regularly come across the use of economic data in these ways.

Economic data can sometimes seem daunting, though perhaps less so if you think of them as creating pictures of economic relationships with real-world observations. In fact, you will soon recognise that we commonly present and use data in a relatively limited number of ways, albeit in many different contexts. You will find that if you are new to the subject, you will very quickly become familiar with these techniques. To help you check your understanding of them, you should visit Web Appendix 1.1 in the Chapter 1 resources of the book's website.[2]

[2]This text does not use calculus. If, however, your course does, you will find some optional Web appendices on the student website or in MyLab Economics, some of which use differentiation to explore some of the theories we cover. The rules of differentiation and their use to find maximum and minimum points are given in Web Appendix 1.2.

Nominal and real data

When using data, one particularly important distinction in economics is that between **nominal** and **real figures.** This distinction applies to all variables measured in monetary terms, such as pounds (£), dollars ($) or euros (€). It also applies to interest rates and exchange rates. It is our fifth threshold concept.

Nominal figures are measured in the actual prices prevailing at the time of measurement. In contrast, real figures correct for changes in prices and so for inflation. But, why is this distinction so important?

KEY IDEA 6 TC 5

The distinction between nominal and real figures. **Nominal figures are those using current prices, interest rates, etc. Real figures are figures corrected for inflation. This distinction is so important in assessing economic data that it is another of our Threshold Concepts.**

Let us consider two applications.

We begin with the concept of **gross domestic product** (**GDP**). This is the value of output produced within a country, typically over a 12-month period. Therefore, it is a measure of a country's national income from production. **Nominal GDP**, sometimes called 'money GDP', measures GDP in the prices of the time (also known as 'current prices'). So, for example, nominal GDP in 2021 would be the value of a country's output at 2021 prices; and nominal GDP in 1999, say, would be in 1999 prices. By simply comparing the figures for the two years, no account is being made for the effect of inflation. If prices had risen between 1999 and 2021, as indeed they had, growth in output would appear to have been higher than it actually was.

In 2021 the nominal (or actual) value of GDP in the UK was £2.3 trillion. It was £1.0 trillion in 1999.

Definitions

Gross domestic product (GDP): The value of output produced within a country, typically over a 12-month period.

Nominal GDP: GDP measured in current prices. These figures take no account of the effect of inflation.

Real GDP: GDP measured in constant prices that ruled in a chosen base year, such as 2000 or 2015. These figures *do* take account of the effect of inflation. When inflation is positive, real GDP figures will grow more slowly than nominal GDP figures.

If we are to make a sensible comparison of national income in these two years, we must take inflation into account. If the 2.3 times increase in national income were merely the result of a 2.3 times increase in the general price level for goods and services, then there would be no *real* increase in national income.

While *nominal* GDP figures take no account of the effect of inflation, **real GDP** figures do. They measure GDP in the prices that ruled in some particular year – the *base year*. Thus, we could measure each year's GDP in, say, 2015 prices (known as 'GDP at constant 2015 prices'). This would enable us to see how much *real* GDP had changed from one period to another by eliminating increases (or decreases) in money GDP due to increases (or decreases) in prices. In real terms GDP in 2021 was only 1.4 times higher than in 1999, not 2.3 times.

Pause for thought

If general price levels are falling, which will rise more quickly: nominal GDP or real GDP?

Real GDP figures therefore allow inferences to be drawn about the percentage (or proportionate) change in the volume of production over time, such as those from one year to the next. These changes are therefore used in measuring economic growth and are reflected in an economy's business cycle. (see Box 1.2). When the rate of inflation is positive, real GDP figures grow more slowly than nominal GDP figures.

The second application of the distinction between nominal and real is when we are comparing the price changes of individual goods to the average price change of all goods. For example, if the prices of bread or houses have risen more than prices in general (i.e. by more than the rate of inflation), we say that their *real* price has risen. If, however, their price has risen by less than the rate of inflation, we say that their real price has fallen.

In both these cases, the transformation from nominal to real figures involves dividing or 'deflating' nominal figures by some general measure of prices. Box 1.3 illustrates this process for the case of UK house prices. The importance of the distinction between nominal and real GDP is discussed further in Chapter 9 (see page 270).

Pause for thought

If the nominal price of a product rises, can its real price fall?

BOX 1.3 NOMINAL AND REAL HOUSE PRICES

Going through the roof

The distinction between nominal and real values is a threshold concept. It is fundamental to assessing economic data. For example, when looking at patterns in the prices of individual goods or services it may be that what we observe is simply reflecting general price patterns. In this case there is no real or relative price change. But, if individual prices are diverging from general price patterns, then a real or relative price change is occurring.

In the UK, house prices are rarely out of the news. This is not only because nominal (actual) prices are increasing over the longer term, but, more significantly, because they are increasing in real terms too. In other words, house prices are increasing *relative* to general prices.

We can quantify the nominal increase in house prices quite straightforwardly. The average UK house price in January 1970 was a little over £3900. By January 2022, as shown in the table, it had risen to around £274 000, an increase of nearly 6900 per cent.

Average UK house price and RPI (January)

	1970	1980	1990	2000	2010	2022
House price (£)	3 920	19 273	58 250	84 620	167 469	273 762
RPI (Jan 1987 = 100)	17.9	62.2	119.5	166.6	217.9	317.7
House price at constant Jan 1987 consumer prices (£)	21 929	30 995	47 445	50 792	76 856	86 170

Sources: *Consumer Price Inflation time series* dataset and *UK House Price Index* (ONS)

But, what about the *real* increase in average UK house prices? To calculate this, we need to compare the increase in house prices with general prices. One index of general prices is the Retail Price Index (RPI).[1] This index measures the cost of a representative basket of consumer goods and services. Over the same period the RPI increased by 1675 per cent. While substantial in its own right, it does mean that house prices have increased more rapidly than consumer prices.

Another way of showing the real increase in house prices is to eliminate the increase in consumer prices from the actual (nominal) house price figures. What is left is the increase in house prices relative to consumer prices. To do this we estimate house prices as if consumer prices had remained at levels at a chosen point in time. In the table we assume that this is January 1987: the RPI equals 100 in January 1987. We then divide the nominal house price figures by the RPI figures before multiplying by 100. This creates a series of average UK house prices at constant January 1987 consumer prices.

The result of this process is illustrated in the final row of the table which contains our real figures. It shows that in real terms the average UK house price increased by around 290 per cent between 1970 and 2022: i.e. the average real house price was 3.9 times more expensive in 2022 than in 1970. The key point, therefore, is that general price inflation cannot explain all the long-term growth seen in average house prices.

The chart plots the average nominal and real UK house price each month from January 1970. The real series shows not only a long-term relative increase in house prices but also *cycles* in the relative price of houses. Hence, a second important finding is the volatility in real house prices. These findings are crucial

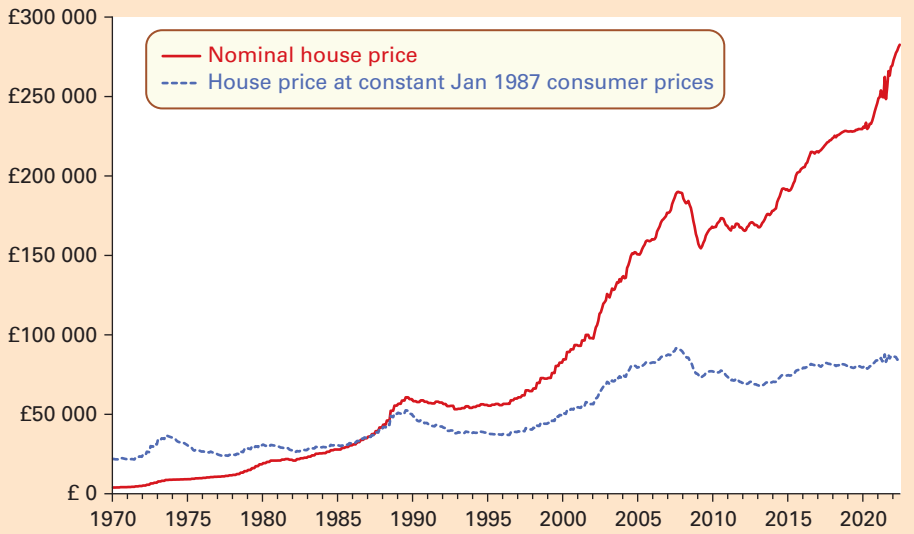
[1] We use RPI figures as the data go back further than for other measures. For other price indices see Section 12.2 and Figure 12.6.

Recap

1. The production possibility curve shows the possible combinations of two goods that a country can produce in a given period of time. Assuming that the country is already producing on the curve, the production of more of one good will involve producing less of the other. This opportunity cost is illustrated by the slope of the curve.
2. If the economy is producing within the production possibility curve as a result of idle resources or inefficiency, it can produce more of both goods by taking up this slack. In the longer term it can only produce more of both by shifting the curve outwards through investment, technological progress, etc.
3. The circular flow of goods and incomes shows the interrelationships between firms and households in a money economy. Firms and households come together in markets. In goods markets, firms supply goods and households demand goods. In the process, money flows from households to firms in return for the goods and services that the firms supply. In factor markets, firms demand factors of production and households supply them. In the process, money flows from firms to households in return for the services of the factors that households supply.
4. When using economic data an important distinction is that between nominal and real figures. This distinction applies to all variables measured in monetary terms. Nominal figures are measured in the actual prices prevailing at the time of measurement, whereas real figures correct for changes in the general price level. In the presence of inflation, real figures increase less quickly than nominal figures.

CASE STUDIES & APPLICATIONS

Average UK house prices



Sources: *Consumer Price Inflation time series dataset* and *UK House Price Index: reports* (Office for National Statistics)

for informing research into the determination of UK house prices.

1. *If house prices were to grow at the same rate as consumer prices over the long term, at what rate would real house prices have increased?*
2. *Under what circumstance would nominal house prices increase while real house prices decrease?*
3. *What is the difference between 'cycles in nominal house prices' and 'cycles in real house prices'?*

Begin by extracting data on UK house prices and the Retail Price Index (or an alternative measure of general prices) and replicate the chart in the Box. These data are available from the Office for National Statistics. Write a short explanation of how you created the real house price series before summarising what your chart shows.

Finally, draw a line chart showing the annual percentage changes in nominal and real house prices and again summarise what your chart shows.

1.5 ECONOMIC SYSTEMS

How do countries differ in the way their economies are organised?

KI 1 p 5

All societies are faced with the problem of scarcity. They differ considerably, however, in the way they tackle the problem. One important difference between societies is in the degree of government control of the economy.

At the one extreme lies the completely **planned or command economy**, where all the economic decisions are taken by the government.

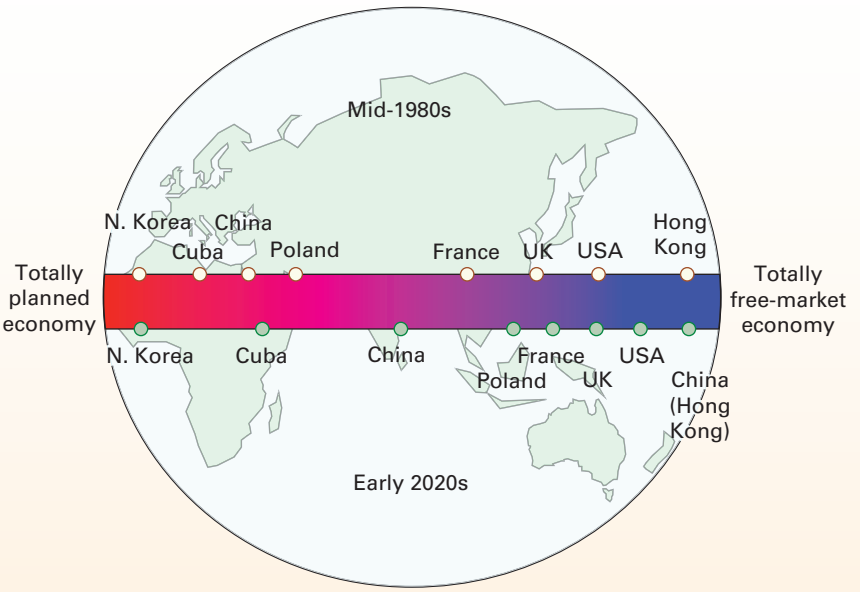
At the other extreme lies the completely **free-market economy**. In this type of economy there is no government intervention at all. All decisions are taken by individuals and firms. Households decide how much labour and other factors to supply, and what goods to consume. Firms decide what goods to produce and what factors to employ. The pattern of production and consumption that results depends on the interactions of all these individual demand and supply decisions.

Definitions

Centrally planned or command economy An economy where all economic decisions are taken by the central authorities.

Free-market economy An economy where all economic decisions are taken by individual households and firms, with no government intervention.

Figure 1.6 Classifying economic systems



In practice, all economies are a mixture of the two and so can be described as **mixed economies**. However, mixed economies can be distinguished by the *degree* of government intervention. For example, the government plays a large role in North Korea, whereas in the USA, the government plays a much smaller role. The mixture of government and the market can be shown by the use of a spectrum diagram such as Figure 1.6. It shows where particular economies of the real world lie along the spectrum between the two extremes.

Pause for thought

How do you think the positions of these eight countries will change over the next decade?

Of course, this analysis presents a very simplified representation of different economies. Mixed economies differ in the *type* of government intervention as well as the degree. For example, governments can intervene through planning, public ownership, regulation, taxes and subsidies, partnership schemes with private industry and so on. Two countries could be in a similar position along the spectrum but have very different types of government intervention.

Nevertheless, this type of analysis is a useful starting point for analysing real-world economies. Also, by looking at the two extremes of a totally planned and a totally free-market economy, this can give us a perspective of the potential advantages and disadvantages of more or less government intervention.

We start by having a brief look at the command economy. Then we will see how a free-market economy operates. We conclude by considering some general reasons for government intervention in market economies. In subsequent chapters we will examine in more detail the various ways in which governments intervene in market economies: and so we will look at the various forms of **mixed market economy**.

The command economy

The command economy is usually associated with a socialist or communist economic system, where land and capital are collectively owned. The state plans the allocation of resources at three important levels.

- It plans the allocation of resources between current consumption and investment for the future. By sacrificing some present consumption and diverting

Definitions

Mixed economy An economy where economic decisions are made partly by the government and partly through the market.

Mixed market economy A market economy where there is some government intervention.

resources into investment, it could increase the economy's growth rate. The amount of resources it chooses to devote to investment will depend on its broad macroeconomic strategy: the importance it attaches to growth as opposed to current consumption.

- At a microeconomic level it plans the output of each industry and firm, the techniques that will be used and the labour and other resources required by each industry and firm.

  In order to ensure that the required inputs are available, the state would probably conduct some form of **input–output analysis**. All industries are seen as users of *inputs* from other industries and as producers of *outputs* for consumers or other industries. For example, the steel industry uses inputs from the coal and iron-ore industries and produces outputs for the vehicle and construction industries. Input–output analysis shows, for each industry, the sources of all its inputs and the destination of all its outputs. The state then attempts to match up the inputs and outputs of each industry so that the planned demand for each industry's product is equal to its planned supply.

TC 4 p 13

- It plans the distribution of outputs between consumers. This will depend on the government's aims. It may distribute goods according to its judgement of people's *needs*; or it may give more to those who produce more, thereby providing an *incentive* for people to work harder.

It may distribute goods and services directly (for example, by a system of rationing); or it may decide the distribution of money incomes and allow individuals to decide how to spend them. If it does the latter, it may still seek to influence the pattern of expenditure by setting appropriate prices: low prices to encourage consumption, and high prices to discourage consumption.

Assessment of the command economy

With central planning, the government could take an overall view of the economy. It could direct the nation's resources in accordance with specific national goals.

High growth rates could be achieved if the government directed large amounts of resources into investment. Unemployment could be largely avoided if the government carefully planned the allocation of labour in accordance with production requirements and labour skills. National income could be distributed more equally or in accordance with needs. The social repercussions of production and consumption (e.g. the effects on the environment) could be taken into account, provided the government was able to predict these effects and chose to take them into account.

In practice, a command economy could achieve these goals only at considerable social and economic cost. The reasons are as follows:

- The larger and more complex the economy, the greater the task of collecting and analysing the information essential to planning, and the more complex the plan. Complicated plans are likely to be costly to administer and involve cumbersome bureaucracy.
- If there is no system of prices, or if prices are set arbitrarily by the state, planning is likely to involve the inefficient use of resources. It is difficult to assess the relative efficiency of two alternative techniques that use different inputs if there is no way in which the value of those inputs can be ascertained. For example, how can a rational decision be made between an oil-fired and a coal-fired furnace if the prices of oil and coal do not reflect their relative scarcity?
- It is difficult to devise appropriate incentives to encourage workers and managers to be more productive without a reduction in quality. For example, if bonuses are given according to the quantity of output produced, a factory might produce shoddy goods, since it can probably produce a larger quantity of goods by cutting quality. To avoid this problem, a large number of officials may have to be employed to check quality.
- Complete state control over resource allocation would involve a considerable loss of individual liberty. Workers would have no choice where to work; consumers would have no choice what to buy.
- The government might enforce its plans even if they were unpopular.
- If production is planned, but consumers are free to spend incomes as they wish, then there will be a problem if the consumer wishes change. Shortages will occur if consumers decide to buy more, and surpluses will occur if they decide to buy less.

Most of these problems were experienced in the former Soviet Union and the other Eastern bloc countries, and were part of the reason for the overthrow of their Communist regimes (see Box 1.4).

Definition

Input–output analysis This involves dividing the economy into sectors where each sector is a user of inputs from and a supplier of outputs to other sectors. The technique examines how these inputs and outputs can be matched to the total resources available in the economy.

Pause for thought

Queues were a common feature of the former Soviet Union. Why do you think they were so commonplace? Is a system of queuing a fair way of allocating scarce goods and resources?

BOX 1.4 COMMAND ECONOMIES

Rise and fall of planning in Russia

Collectivisation and central planning

The Bolsheviks under the leadership of Lenin came to power in Russia with the October Revolution of 1917. Communism was introduced and the market economy abolished. Industries were nationalised; workers were told what jobs to do; food was taken from peasants to feed the towns; workers were allocated goods from distribution depots.

With the ending of the civil war in 1921, the economy was in bad shape and Lenin embarked on the New Economic Policy. This involved a return to the use of markets. Smaller businesses were returned to private hands and peasants were able to sell their crops. The economy began to recover; however, Lenin died in 1924 and Stalin came to power.

The Russian economy underwent a radical transformation from 1928 onwards. The key features of the Stalinist approach were collectivisation, industrialisation and central planning. Peasant farms were abolished and replaced by large-scale collective farms where land was collectively owned and worked, and by state farms, owned by the state and run by managers. This caused disruption and famine, with peasants slaughtering their animals rather than giving them up. However, in the longer term more food was produced. Both collective and state farms were given quotas of output that they were supposed to deliver, for which the state would pay a fixed price.

Alongside the agricultural reforms a drive to industrialisation took place and a vast planning apparatus was developed. At the top was *Gosplan,* the central planning agency. This prepared five-year plans, which specified the general direction in which the economy was to move, and annual plans, which gave details of what was to be produced and with what resources for some 200 or so key products. The system operated without either the price mechanism or the profit motive, although incentives existed with bonuses paid to managers and workers if targets were achieved.

Stalin died in 1953, but the planning system remained largely unchanged throughout the Soviet Union until the late 1980s. Initially, high growth rates had been achieved, though at a cost of low efficiency. Poor flows of information led to inconsistencies in the plans. Targets were often unrealistic, and as a result there were frequent shortages and sometimes surpluses. There was little product innovation and goods were frequently of poor quality. A large 'underground economy' flourished in which goods were sold on the illegal market and in which people did second 'unofficial' jobs.

Moves to the market

By the time Gorbachev came to power in 1985 many people were pressing for economic reform. Gorbachev responded with his policy of *perestroika* (economic reconstruction), which involved managers preparing their own plans and managers and workers being rewarded for becoming more efficient. Under the new system, one-person businesses and larger co-operatives were allowed, while the price mechanism was re-introduced with the state raising prices if there were substantial shortages.

These reforms, however, did not halt the economic decline. Managers resented the extra responsibilities and people were unclear as to what to expect from the state. Queues lengthened in the shops and people became disillusioned with *perestroika*.

Communism fell apart in 1989 and both the Soviet Union and the system of central planning came to an end. Russia embarked upon a radical programme of market reforms in which competition and enterprise were intended to replace state central planning (see Case Study 1.7 in MyLab Economics.)

Initially, the disruption of the move to the market led to a sharp decline in the Russian economy. Output fell by an average of 5.5 per cent per annum between 1993 and 1998. However, this was followed by a period of rapid economic growth, which averaged 7 per cent from 1999 to 2008.

The Russian economy contracted by nearly 8 per cent in the 2009 recession. Although this was followed by growth

The free-market economy

Free decision making by individuals

The free-market economy is usually associated with a pure capitalist system, where land and capital are privately owned. All economic decisions are made by households and firms, which are assumed to act in their own self-interest. The following assumptions are usually made:

- Firms seek to maximise profits.
- Consumers seek to get the best value for money from their purchases.
- Workers seek to maximise their wages relative to the human cost of working in a particular job.

It is also assumed that individuals are free to make their own economic choices: consumers are free to decide what to buy with their incomes; workers are free to choose where and how much to work; firms are free to choose what to sell and what production methods to use.

The resulting supply and demand decisions of firms and households are transmitted to each other through their effect on *prices*.

The price mechanism

The **price mechanism** works as follows. Prices respond to *shortages* and *surpluses*. Shortages cause prices to rise. Surpluses cause prices to fall.

If consumers decide they want more of a good (or if producers decide to cut back supply), demand will

Definition

Price mechanism The system in a market economy whereby price changes that occur in response to changes in demand and supply have the effect of making demand equal to supply.

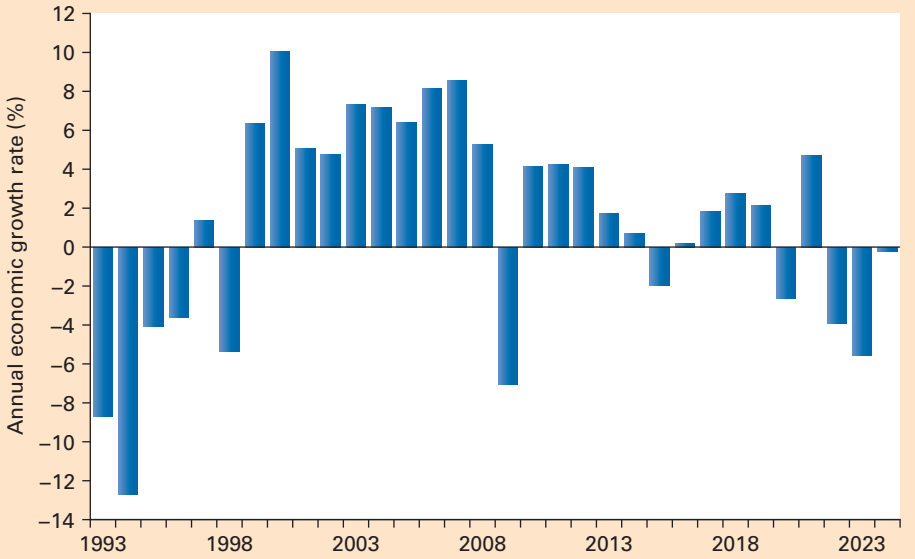
CASE STUDIES & APPLICATIONS

averaging nearly 4.5 per cent between 2010 and 2012, growth then weakened. Many commentators pointed to decades of underinvestment in industry and in road and rail infrastructure, corruption, disillusionment and continuing political uncertainty as root causes of this sluggish growth. Inequality grew as the rich 'oligarchs' got richer, with little trickling down to the poor.

Then from 2014, the economy was further dampened by Western economic sanctions in response to Russia's annexation of Crimea and the continuing conflict in Eastern Ukraine. The Russian economy contracted by 2.5 per cent in 2015 and by a further 0.2 per cent in 2016. Then, with the invasion of Ukraine in 2022 and the imposition of stronger western sanctions, the economy went into steep recession.

Research the process to a more market-based economy that has taken place in China. Summarise the process in a short briefing note including an economic growth chart similar to the one for Russia in this box.

Economic growth in Russia



Note: Figures from 2022 based on forecasts

Source: Data drawn from *World Economic Outlook Database* (IMF, October 2022) and various forecasts

exceed supply. The resulting *shortage* will encourage sellers to *raise* the price of the good. This will act as an incentive to producers to supply more, since production will now be more profitable. At the same time, it will discourage consumers from buying so much. *The price will continue rising until the shortage has been eliminated.*

If, on the other hand, consumers decide they want less of a good (or if producers decide to produce more), supply will exceed demand. The resulting *surplus* will encourage sellers to *reduce* the price of the good. This will act as a disincentive to producers, who will supply less, since production will now be less profitable. It will encourage consumers to buy more. *The price will continue falling until the surplus has been eliminated.*

This price, where demand equals supply, is called the **equilibrium price**. By **equilibrium** we mean a point of balance or a point of rest: in other words, a point towards which there is a tendency to move.

TC 4 p 13

The same analysis can be applied to factor markets. If the demand for a particular type of labour exceeds its supply, the resulting shortage will drive up the

Definitions

Equilibrium price The price where the quantity demanded equals the quantity supplied: the price where there is no shortage or surplus.

Equilibrium A position of balance. A position from which there is no inherent tendency to move away.

wage rate (i.e. the price of labour) as employers compete with each other for labour. The rise in the wage rate will have the effect of curbing firms' demand for that type of labour and encouraging more workers to take up that type of job. Wages will continue rising until demand equals supply: until the shortage is eliminated.

Likewise, if there is a surplus of a particular type of labour, the wage will fall until demand equals supply. As with price, the wage rate where the demand for labour equals the supply is known as the *equilibrium* wage rate.

The response of demand and supply to changes in price illustrates a very important feature of how economies work: *people respond to incentives*. It is important, therefore, that incentives are appropriate and have the desired effect. This is our sixth threshold concept.

KEY IDEA 7

People respond to incentives. **It is important, therefore, that incentives are appropriate and have the desired effect. This is another of our threshold concepts. For details, see MyLab Economics.**

TC 6

The effect of changes in demand and supply

How will the price mechanism respond to changes in consumer demand or producer supply? After all, patterns of consumer demand will change over time. For example, people may decide they want more ethically sourced food and drink compared with other food and drink products. Likewise, the pattern of supply also changes. For example, changes in technology may allow the mass production of microchips at lower cost, while the production of hand-built furniture becomes relatively expensive.

In all cases of changes in demand and supply, the resulting changes in *price* act as both *signals* and *incentives*.

A change in demand. A rise in demand is signalled by a rise in price. This then acts as an incentive for firms to produce more of the good: the quantity supplied rises. Firms divert resources from goods with lower prices relative to costs (and hence lower profits) to those goods that are more profitable.

TC 6 p 24

A fall in demand is signalled by a fall in price. This then acts as an incentive for firms to produce less: such goods are now less profitable to produce. Thus the quantity supplied falls.

A change in supply. A rise in supply is signalled by a fall in price. This then acts as an incentive for consumers to buy more: the quantity demanded rises. A fall in supply is signalled by a rise in price. This then acts as an incentive for consumers to buy less: the quantity demanded falls.

KEY IDEA 8

Changes in demand or supply cause markets to adjust. **Whenever such changes occur, the resulting 'disequilibrium' will bring an automatic change in prices, thereby restoring equilibrium (i.e. a balance of demand and supply).**

The interdependence of markets

The interdependence of goods and factor markets. A rise in demand for a good will raise its price and profitability. Firms will respond by supplying more. But to do this they will need more inputs. Thus the demand for the inputs will rise, which in turn will raise the price of the inputs. The suppliers of inputs will respond to this incentive by supplying more. This can be summarised as follows:

TC 6 p 24

1. Goods market
   - Demand for the good rises.
   - This creates a shortage.
   - This causes the price of the good to rise.
   - This eliminates the shortage by choking off some of the demand and encouraging firms to produce more.
2. Factor market
   - The increased supply of the good causes an increase in the demand for factors of production (i.e. inputs) used in making it.
   - This causes a shortage of those inputs.
   - This causes their prices to rise.
   - This eliminates their shortage by choking off some of the demand and encouraging the suppliers of inputs to supply more.

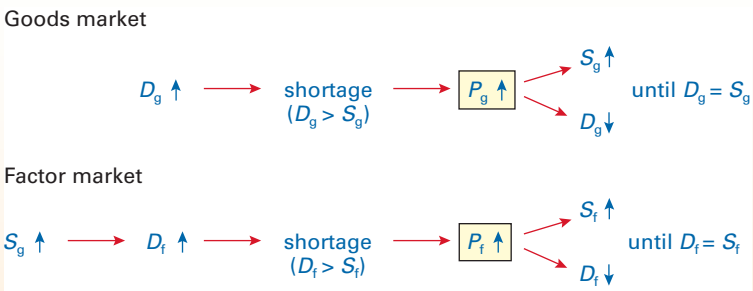
KI 8 p 24

Goods markets thus affect factor markets.

It is common in economics to summarise an argument like this by using symbols. It is a form of shorthand. Figure 1.7 summarises this particular sequence of events.

Interdependence exists in the other direction too: factor markets affect goods markets. For example, the discovery of raw materials will lower their price. This will lower the production costs of firms using these raw materials and increase the supply of the finished good. The resulting surplus will lower the

Figure 1.7 The price mechanism: the effect of a rise in demand



price of the good, which will encourage consumers to buy more.

The interdependence of different goods markets. A rise in the price of one good will encourage consumers to buy alternatives. This will drive up the price of alternatives. This in turn will encourage producers to supply more of the alternatives.

Pause for thought

Summarise this last paragraph using symbols like those in Figure 1.7.

TC 6 p 24

Interdependence and the public interest. Even though all individuals are merely looking to their own self-interest in the free-market economy, they are in fact being encouraged to respond to the wishes of others through the incentive of the price mechanism. For example, if consumers want more of a product, firms will supply more – not out of the goodness of their hearts, but because it is profitable to do so. It is often claimed that this is a major advantage of a free-market economy. We will be examining this claim in subsequent chapters.

Competitive markets

In the next chapter we will examine the working of the price mechanism in more detail. We will look first at demand, then at supply, and then we will put the two together to look at the determination of price.

The markets we will be examining are highly competitive markets, with many firms competing against each other. In economics we call this **perfect competition**. This is where consumers and producers are too numerous to have any control over prices: they are **price takers**.

In the case of consumers, this means that they have to accept the prices as given for the things that they buy. On most occasions this is true. For example, when you get to the supermarket checkout you cannot start haggling with the checkout operator over the price of a can of beans or a bar of chocolate.

In the case of firms, perfect competition means that producers are too small and face too much competition from other firms to be able to raise prices. Take the case of farmers selling wheat. They have to sell it at the current market price. If individually they try to sell at a higher price, no one will buy, since purchasers of wheat (e.g. flour millers) can get all the wheat they want at the market price.

Of course, many firms *do* have the power to choose their prices. This does not mean that they can simply charge whatever they like. They will still have to take account of overall consumer demand and their competitors' prices. Ford, when setting the price of its Focus cars, will have to ensure that they remain competitive with Astras, Golfs, Peugeot 308s, etc. Nevertheless, most firms have some flexibility in setting their prices: they have a degree of 'market power'.

If this is the case, then why do we study *perfect* markets, where firms are price takers? One reason is that they provide a useful approximation to the real world and give us many insights into how a market

TC 4 p 13

Definitions

Perfect competition (preliminary definition) A situation where the consumers and producers of a product are price takers. (There are other features of a perfectly competitive market; these are examined in Chapter 6.)

Price taker A person or firm with no power to be able to influence the market price.

economy works. Many markets do function very similarly to the markets we shall be describing.

Another is that perfect markets provide an ideal against which to compare the real world. It is often argued that perfect markets benefit the consumer, whereas markets dominated by big business may operate against the consumer's interests. For example, the consumer may end up paying higher prices in a market dominated by just a few firms than in one operating under perfect competition.

The mixed market economy

A 'mixed system' is one containing elements of markets and government control. Most economies today can be described as 'mixed market economies', where markets are the predominant means by which a society's scarce resources are allocated, but where there is some government intervention. However, both the *degree* to which governments intervene and the *types* of intervention vary from country to country and from government to government within countries.

Government intervention generates considerable debate among not only economists, but also policy makers and the wider public too. Such debates intensified in the wake of the financial crisis of the late 2000s, the subsequent economic downturn and so-called austerity measures taken by many governments to reduce the size of their deficits (the excess of government spending over receipts). There was further debate during the COVID-19 pandemic over how much the government should support workers and businesses, and again in 2022 over what support should be provided to people, especially those on low incomes, as energy and food prices soared.

Subsequent chapters will consider government intervention in market economies. But we conclude this chapter by outlining three broad *economic* reasons for intervention. These can, in turn, be used as a framework to help analyse the effects of government interventions.

Definitions

Allocative role Interventions by government to affect the allocation of resources in consumption and/or production.

Distributive role Interventions by government to affect the distribution of resources such as the distribution of incomes.

Macroeconomic role Interventions by government either to stabilise the economy in the short term or to promote longer-term economic growth.

First, a government may intervene to affect the allocation of resources: an **allocative role**. More specifically, it can attempt to affect levels of consumption and/or production. For instance, it may choose to provide compulsory primary and secondary education in the belief that in not doing so too little would be provided and/or too little would be consumed. On the other hand, it may choose to levy taxes on certain products, such as cigarettes and alcohol, to raise their price and thereby discourage their consumption.

TC 6 p 24

Second, the government may choose to affect the distribution of resources and/or the resulting distribution of income, wealth and welfare: a **distributive role**. The income tax system, for instance, can be designed to create a more equal distribution of income, while the subsidising or provision of services, such as libraries and parks, is a means of granting more equal access to society's scarce resources.

Third, the government may intervene to reduce the volatility of the macroeconomy and/or to help promote the long-term growth of the economy: a **macroeconomic role**. This role, particularly the stabilisation function, is keenly debated by economists. However, many economists recognise the role that governments can play in helping to advance longer-term growth by providing, for instance, infrastructure and education and training.

TC 3 p 10

Pause for thought

How can government regulation be a tool for achieving the three economic roles of government?

The fact that government intervention can, at least in principle, correct market failures is the seventh of our fifteen threshold concepts.

KEY IDEA 9 / TC 7

Government intervention may be able to rectify various failings of the market. Government intervention in the market can be used to achieve various economic objectives that may not be best achieved by the market. Governments are not perfect, however, and their actions may bring adverse as well as beneficial consequences.

Pause for thought

Why do governments on the political right tend to intervene less in markets than governments on the political left? Does this mean that whether something is an economic 'problem' depends on your perspective?

Recap

1. The economic systems of different countries vary according to the extent to which they rely on the market or the government to allocate resources.
2. At the one extreme, in a command economy, the state makes all the economic decisions. It plans how many resources to allocate for present consumption and how many for investment for future output. It plans the output of each industry, the methods of production it will use and the amount of resources it will be allocated. It plans the distribution of output between consumers.
3. A command economy has the advantage of being able to address directly various national economic goals, such as rapid growth and the avoidance of unemployment and inequality. A command economy, however, is likely to be inefficient: a large bureaucracy will be needed to collect and process information; prices and the choice of production methods are likely to be arbitrary; incentives may be inappropriate; shortages and surpluses may result.
4. At the other extreme is the free-market economy. In this economy, decisions are made by the interaction of demand and supply. Price changes act as the mechanism whereby demand and supply are balanced. If there is a shortage of a product, its price will rise until the shortage is eliminated. If there is a surplus, its price will fall until that is eliminated.
5. Perfect markets are markets where both producers and consumers are price takers.
6. A mixed market economy is a predominantly market-based economic system with some government intervention. We can identify three broad roles for government intervention: allocative, distributive and macroeconomic.

QUESTIONS

1. Imagine that you won millions of pounds on the National Lottery. Would your 'economic problem' be solved?
2. Would redistributing incomes from the rich to the poor reduce the overall problem of scarcity?
3. In what way does specialisation reduce the problem of scarcity?
4. Which of the following are macroeconomic issues, which are microeconomic ones, and which could be either depending on the context?
   a. Inflation.
   b. Low wages in certain service industries.
   c. The rate of exchange between the pound and the euro.
   d. Why the price of cabbages fluctuates more than that of cars.
   e. The rate of economic growth this year compared with last year.
   f. The decline of traditional manufacturing industries.
   g. A decline in house prices
   h. A reduction in the supply of credit by financial institutions
5. Assume that in a household one parent currently works full time and the other stays at home to look after the family. How would you set about identifying and calculating the opportunity costs of the second parent now taking a full-time job? How would such calculations be relevant in deciding whether it is worth taking that job?
6. When you made the decision to study economics, was it a 'rational' decision (albeit based on the limited information you had available at the time)? What additional information would you like to have had in order to ensure that your decision was the right one?
7. Assume you are looking for a job and are offered two. One is more unpleasant to do, but pays more.
   How would you make a rational choice of which of the two jobs to accept?
8. Imagine that a country can produce just two things: goods and services. Assume that over a given time period it could produce any of the following combinations:

Units of goods	0	10	20	30	40	50	60	70	80	90	100
Units of services	80	79	77	74	70	65	58	48	35	19	0

   a. Draw the country's production possibility curve.
   b. Assuming that the country is currently producing 40 units of goods and 70 units of services, what is the opportunity cost of producing another 10 units of goods?
   c. Explain how the figures illustrate the principle of increasing opportunity cost.
   d. Now assume that technical progress leads to a 10 per cent increase in the output of goods for any given amount of resources. Draw the new production possibility curve. How has the opportunity cost of producing extra units of services altered?
9. Under what circumstances would the production possibility curve be (a) a straight line, (b) bowed in towards the origin? Are these circumstances ever likely?
   a. If a product's price has been rising over time, how can we explain the possibility that its real price has fallen?
   b. In 2008, current-price GDP in the UK rose by 2.3 per cent. However, the UK economy contracted by 0.5 per cent. How can this apparent contradiction be explained?
10. Using a chart similar to that in Figure 1.7, trace through the following effects: (a) a fall in demand for a good; (b) an increased supply of a factor of production. (In the case of (b) you will need to consider the factor market first, and then the goods market.)
11. Draw up a list of policy measures that may help government to affect: (i) the allocation of resources; (ii) the distribution of resources; (iii) macroeconomic stability; (iv) long-term economic growth. Is it possible that the policies you have identified in each case could conflict with each other?

Pearson | MyLab | Economics

Log on to MyLab Economics and complete the studyplan exercises for this chapter to see how much you have learnt and where you need to revise most. Make sure you access all the supporting textbook resources, including the online workbook, news-blog, audio animations, guided solutions and ebook.

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

1.1 **Scarcity and abundance.** If scarcity is the central economic problem, is anything truly abundant?

1.2 **Global economics.** This examines how macroeconomics and microeconomics apply at the global level and identifies some key issues.

1.3 **Buddhist economics.** A different perspective on economic problems and economic activity.

1.4 **Green economics.** This examines some of the environmental costs that society faces today. It also looks at the role of economics in analysing these costs and how the problems can be tackled.

1.5 **Positive and normative statements.** A crucial distinction when considering matters of economic policy.

1.6 **Adam Smith (1723–90).** Smith, the founder of modern economics, argued that markets act like an invisible hand guiding production and consumption.

1.7 **Free-market medicine in Russia.** This extends the analysis of Box 1.4 by examining the operation of the market economy in Russia and the successes and difficulties in moving from a planned to a market economy.

1.8 **Alternative measures of well-being.** This case study takes a preliminary look at how we measure the well-being of society. Should we use output (GDP) per head or some other measure?

WEBSITES RELEVANT TO PART A

Numbers and sections refer to websites listed in the Websites Appendix at the end of the book and hotlinked from this book's website at www.pearsoned.co.uk/sloman.

- For news articles relevant to this Introduction, search for the *Sloman Economics News* site or follow the News Items link from the student website or MyLab Economics.
- For general economics news sources, see websites in section A of the Web Appendix at the end of the book, and particularly A1–9, 24, 25, 35, 36. See also A39–44 for links to newspapers worldwide.
- For sources of economic data, see sites in section B and particularly B1–6, 29, 31, 34, 35, 38, 47.
- For general sites for students of economics, see sites in section C and particularly C1–10.
- For sites giving links to relevant economics websites, organised by topic, see sites I7, 11, 12, 16, 18.
- For news on the Russian economy (Box 1.4 and Case Study 1.7 on the student website), see sites A14, 15; B31

WEB APPENDICES

1.1 **Some techniques of economic analysis.** This appendix explains how economists use graphs, tables and simple equations to represent economic data.

1.2 **Elementary differentiation.** This gives the rules for simple differentiation. Although this book does not use calculus (of which differentiation is part), the techniques are widely used by economists and you will need to use calculus if you progress to more advanced economic courses.

Part B

Microeconomics

Chapter 2

Markets, demand and supply

In this section of the text we focus on microeconomics. Despite being 'small economics' – in other words, the economics of the individual parts of the economy, rather than the economy as a whole – it is still concerned with many of the big issues of today.

We will study why the pattern of production and consumption changes over time; why some people are rich and others poor; why our lives seem to be dominated by market forces beyond our control. We will look at the world of big business at one extreme and highly competitive markets at the other. We will look at many of the seemingly intractable problems we face: from the growing problem of pollution, to our limited power as consumers, to the widening inequality of incomes in society.

In this and the next chapter, we will look at how markets work. We will examine what determines how much of any product gets produced and sold, and why some goods rise in price, whereas others fall. In the process we will be looking at one of the most important theories in the whole of economics: the theory of supply and demand.

After studying this chapter, you should be able to answer the following questions:

- How do markets operate?
- What determines the amount that consumers buy of a product?
- What determines how much producers supply of a product?
- How are market prices determined and when are they likely to rise or fall?
- How do markets respond to changes in demand or supply?
- What are the major strengths and weaknesses of a free-market economy?

2.1 DEMAND

How much will people buy of any item?

The relationship between demand and price

The headlines announce, 'Major crop failures in Brazil and East Africa: coffee prices soar'. Shortly afterwards you find that coffee prices have doubled in the shops. What do you do? Presumably you will cut back on the amount of coffee you drink. Perhaps you will reduce it from, say, six cups per day to two. Perhaps you will give up drinking coffee altogether.

This is simply an illustration of the general relationship between price and consumption: *when the price of a good rises, the quantity demanded will fall*. This relationship is known as the **law of demand**. There are two reasons for this law.

TC 5 p 17

- People will feel poorer. They will not be able to afford to buy so much of the good with their money: the purchasing power of their income has fallen. Hence their *real income* has fallen. This is called the **income effect** of a price rise.
- The good will now cost more relative to other goods. People will thus switch to alternative or 'substitute' goods. This is called the **substitution effect** of a price rise.

Similarly, when the price of a good falls, the quantity demanded will rise. People can afford to buy more (the income effect), and some will switch to consuming this good from alternative goods (the substitution effect).

Therefore, returning to our example of the increase in the price of coffee, we will not be able to afford to buy as much as before, and we will probably drink more tea, cocoa, fruit juices or even water instead.

A word of warning: be careful about the meaning of the words **quantity demanded**. They refer to the amount consumers are willing and able to purchase at a given price over a given period (e.g. a week, or a month, or a year). They do *not* refer to what people would simply *like* to consume. You might like to own a luxury yacht, but your demand for luxury yachts will almost certainly be zero at the current price.

The demand curve

Consider the hypothetical data in Table 2.1. The table shows how many kilos of potatoes per month would be purchased at various prices.

Columns (2) and (3) show the **demand schedules** for two individuals, Dean and John. Column (4), by contrast, shows the total **market demand schedule**. This is the total demand by all consumers. To obtain the market demand schedule for potatoes, we simply add up the quantities demanded at each price by *all* consumers: i.e. Dean, John and everyone else who demands potatoes. Notice that we are talking about demand *over a period of time* (not at a *point* in time). Thus we would talk about daily demand, or weekly demand, or annual demand or whatever.

Table 2.1 The demand for potatoes (monthly)

	Price (pence per kilo) (1)	Dean's demand (kilos) (2)	John's demand (kilos) (3)	Total market demand (tonnes: 000s) (4)
A	20	28	16	700
B	40	15	11	500
C	60	5	9	350
D	80	1	7	200
E	100	0	6	100

The demand schedule can be represented graphically as a **demand curve**. Figure 2.1 shows the market

Definitions

Law of demand The quantity of a good demanded per period of time will fall as price rises and will rise as price falls, other things being equal.

Income effect The effect of a change in price on quantity demanded arising from the consumer becoming better or worse off as a result of the price change.

Substitution effect The effect of a change in price on quantity demanded arising from the consumer switching to or from alternative (substitute) products.

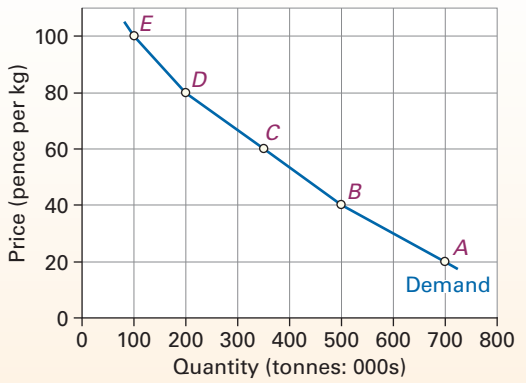
Quantity demanded The amount of a good that a consumer is willing and able to buy at a given price over a given period of time.

Demand schedule for an individual A table showing the different quantities of a good that a person is willing and able to buy at various prices over a given period of time.

Market demand schedule A table showing the different total quantities of a good that consumers are willing and able to buy at various prices over a given period of time.

Demand curve A graph showing the relationship between the price of a good and the quantity of the good demanded over a given time period. Price is measured on the vertical axis; quantity demanded is measured on the horizontal axis. A demand curve can be for an individual consumer or group of consumers, or more usually for the whole market.

Figure 2.1 Market demand curve for potatoes (monthly)



demand curve for potatoes corresponding to the schedule in Table 2.1. The price of potatoes is plotted on the vertical axis. The quantity demanded is plotted on the horizontal axis.

Point *E* shows that at a price of 100p per kilo, 100 000 tonnes of potatoes are demanded each month. When the price falls to 80p we move down the curve to point *D*. This shows that the quantity demanded has now risen to 200 000 tonnes per month. Similarly, if the price falls to 60p we move down the curve again to point *C*: 350 000 tonnes are now demanded. The five points on the graph (*A*–*E*) correspond to the figures in columns (1) and (4) of Table 2.1. The graph also enables us to read off the likely quantities demanded at prices other than those in the table.

A demand curve could also be drawn for an individual consumer. Like market demand curves, individuals' demand curves generally slope downwards from left to right: the lower the price of a product, the more a person is likely to buy.

Two points should be noted at this stage.

- In textbooks, demand curves (and other curves too) are only occasionally used to plot specific data. More frequently they are used to illustrate general theoretical arguments. In such cases, the axes will simply be price and quantity, with the units unspecified.
- The term demand 'curve' is used even when the graph is a straight line! In fact, when using demand curves to illustrate arguments we frequently draw them as straight lines – it's simpler.

TC4 p 13

Other determinants of demand

Price is not the only factor that determines how much of a good people will buy. Demand is also affected by the following:

Tastes. The more desirable people find the good, the more they will demand. Tastes are affected by advertising, by fashion, by observing other consumers, by considerations of health and by the experiences from consuming the good on previous occasions. During the 2020 lockdowns, the demand for face-coverings and hand sanitiser increased significantly.

The number and price of substitute goods (i.e. competitive goods). The higher the price of **substitute goods**, the higher will be the demand for this good as people switch from the substitutes. For example, the demand for coffee will depend on the price of tea. If tea goes up in price, the demand for coffee will rise.

The number and price of complementary goods. **Complementary goods** are those that are consumed together: cars and petrol, paper and ink cartridges, fish and chips. The higher the price of complementary goods, the fewer of them will be bought and hence the less will be the demand for this good. For example, the demand for games will depend on the price of games consoles, such as the Sony PlayStation® and Microsoft® Xbox. If the price of games consoles comes down, so that more are purchased, the demand for games will rise.

Income. As people's incomes rise, their demand for most goods will rise. Such goods are called **normal goods**. There are exceptions to this general rule, however. As people get richer, they spend less on **inferior goods**, such as supermarkets' 'value lines', and switch to better-quality goods.

Distribution of income. If national income were redistributed from the poor to the rich, the demand for luxury goods would rise. At the same time, as the poor got poorer they might have to turn to buying inferior goods, whose demand would thus rise too.

Expectations of future price changes. If people think that prices are going to rise in the future, they are likely to buy more now before the price goes up.

Definitions

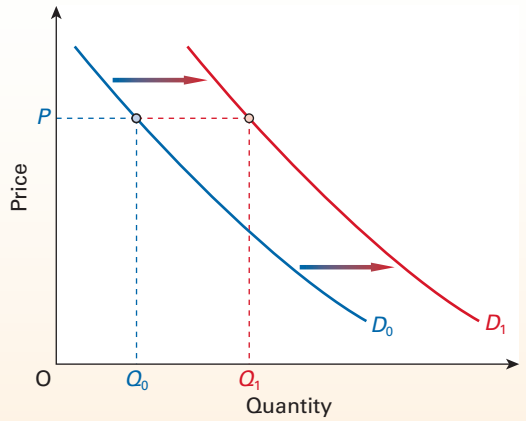
Substitute goods A pair of goods that are considered by consumers to be alternatives to each other. As the price of one goes up, the demand for the other rises.

Complementary goods A pair of goods consumed together. As the price of one goes up, the demand for both goods will fall.

Normal good A good whose demand rises as people's incomes rise.

Inferior good A good whose demand falls as people's incomes rise.

Figure 2.2 An increase in demand



Movements along and shifts in the demand curve

A demand curve is constructed on the assumption that 'other things remain equal' (sometimes known by the Latin term ***ceteris paribus***). In other words, it is assumed that none of the determinants of demand, other than price, changes.[1] The effect of a change in price is then simply illustrated by a movement along the demand curve: for example, from point *B* to point *D* in Figure 2.1 when the price of potatoes rises from 40p to 80p per kilo.

TC 4 p 13

What happens, then, when one of these other determinants does change? The answer is that we have to construct a whole new demand curve: the curve shifts. If a change in one of the other determinants causes demand to rise – say, income rises – the whole curve will shift to the right. This shows that at each price, more will be demanded than before. Thus in Figure 2.2, at a price of P a quantity of Q_0 was originally demanded. But now, after the increase in demand, Q_1 is demanded. (Note that D_1 is not necessarily parallel to D_0.)

If a change in a determinant other than price causes demand to fall, the whole curve will shift to the left.

Pause for thought

The price of cinema tickets rises and yet it is observed that cinema attendance increases. Does this mean that the demand curve for cinema tickets is upward sloping?

To distinguish between shifts in and movements along demand curves, we refer to a shift in demand as a **change in demand**, and a movement along the demand curve as a result of a change in price as a **change in the quantity demanded.**

Definitions

Ceteris paribus Latin for 'other things being equal'. This assumption has to be made when making deductions from theories.

Change in demand The term used for a shift in the demand curve. It occurs when a determinant of demand *other* than price changes.

Change in the quantity demanded The term used for a movement along the demand curve to a new point. It occurs when there is a change in price.

[1] We make this assumption to keep the analysis simple at the outset. We can then drop the assumption by changing things one at a time and seeing what happens.

Recap

1. When the price of a good rises, the quantity demanded per period of time will fall. This is known as the 'law of demand'. It applies both to individuals' demand and to the whole market demand.
2. The law of demand is explained by the income and substitution effects of a price change.
3. The relationship between price and quantity demanded per period of time can be shown in a table (or 'schedule') or as a graph. On the graph, price is plotted on the vertical axis and quantity demanded per period of time on the horizontal axis. The resulting demand curve is downward sloping (negatively sloped).
4. Other determinants of demand include tastes, the number and price of substitute goods, the number and price of complementary goods, income, the distribution of income and expectations of future price changes.
5. If price changes, the effect is shown by a movement along the demand curve. We call this effect 'a change in the quantity demanded'.
6. If any other determinant of demand changes, the whole curve will shift. We call this effect 'a change in demand'. A rightward shift represents an increase in demand; a leftward shift represents a decrease in demand.

2.2 SUPPLY

How much of any item will firms want to produce?

Supply and price

Imagine you are a farmer deciding what to do with your land. Part of your land is in a fertile valley. Part is on a hillside where the soil is poor. Perhaps, then, you will consider growing vegetables in the valley and keeping sheep on the hillside.

TC 2 p 9

Your decision will largely depend on the price that various vegetables will fetch in the market and likewise the price you can expect to get from sheep and wool. As far as the valley is concerned, you will plant the vegetables that give the best return. If, for example, the price of potatoes is high, you will probably use a lot of the valley for growing potatoes. If the price gets higher, you may well use the whole of the valley, perhaps being prepared to run the risk of potato disease. If the price is very high, you may even consider growing potatoes on the hillside, even though the yield per hectare is much lower there. In other words, the higher the price of a particular crop, the more you are likely to grow it in preference to other crops.

This illustrates the general relationship between supply and price: *when the price of a good rises, the quantity supplied will also rise*. There are three reasons for this.

- As firms supply more, they are likely to find that beyond a certain level of output costs rise more and more rapidly. Only if price rises will it be worth producing more and incurring these higher costs.

  In the case of the farm we just considered, once potatoes have to be grown on the hillside the costs of producing them will increase. Also, if the land has to be used more intensively, say by the use of more and more fertilisers, again the costs of producing extra potatoes are likely to rise quite rapidly. It is the same for manufacturers. Beyond a certain level of output, costs are likely to rise rapidly as workers have to be paid overtime and as machines approach capacity working. If higher output involves higher costs of production, producers will need to get a higher price if they are to be persuaded to produce extra output. We consider how costs rise with rises in output in Chapter 5.

TC 2 p 9

- The higher the price of the good, the more profitable it becomes to produce. Firms will thus be encouraged to produce more of it by switching from the production of less profitable goods.
- Given time, if the price of a good remains high, new producers will be encouraged to set up in production. Total market supply thus rises.

The first two determinants affect supply in the short run. The third affects supply in the long run. We distinguish between short-run and long-run supply in Chapter 3 (page 57).

The supply curve

The amount that producers would like to supply at various prices can be shown in a **supply schedule**. Table 2.2 shows a monthly supply schedule for potatoes, both for an individual farmer (farmer X) and for all farmers together (the whole market).

The supply schedule can be represented graphically as a **supply curve**. A supply curve may be an individual firm's supply curve or a market curve (i.e. that of the whole industry).

Figure 2.3 shows the *market* supply curve of potatoes. As with demand curves, price is plotted on the vertical axis and quantity on the horizontal axis. Each of the points *a–e* corresponds to a figure in Table 2.2. For example, a price rise from 60p per kilo to 80p per kilo will cause a movement along the supply curve from point *c* to point *d*: total market supply will rise from 350 000 tonnes per month to 530 000 tonnes per month.

Table 2.2 The supply of potatoes (monthly)

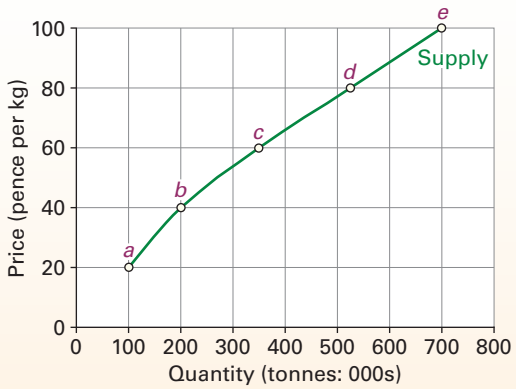
	Price of potatoes (pence per kilo)	Farmer X's supply (tonnes)	Total market supply (tonnes: 000s)
A	20	50	100
B	40	70	200
C	60	100	350
D	80	120	530
E	100	130	700

Definitions

Supply schedule A table showing the different quantities of a good that producers are willing and able to supply at various prices over a given time period. A supply schedule can be for an individual producer or group of producers, or for all producers (the market supply schedule).

Supply curve A graph showing the relationship between the price of a good and the quantity of the good supplied over a specified period of time.

Figure 2.3 Market supply curve of potatoes (monthly)



Not all supply curves will be upward sloping (positively sloped). Sometimes they will be vertical, or horizontal, or even downward sloping. This will depend largely on the time period over which firms' response to price changes is considered. This question is examined in Chapter 3 in the section on the elasticity of supply (Section 3.3) and in more detail in Chapters 5 and 6.

Other determinants of supply

Like demand, supply is not simply determined by price. The other determinants of supply are as follows:

The costs of production. The higher the costs of production, the less profit will be made at any price. As costs rise, firms will cut back on production, probably switching to alternative products whose costs have not risen so much.

The main reasons for a change in costs are:

- Change in input prices: costs of production will rise if wages, raw material prices, rents, interest rates or other input prices rise.
- Change in technology: technological advances can fundamentally alter the costs of production. The microchip revolution, for example, has changed production methods and information handling in virtually every industry in the world.
- Organisational changes: various cost savings can be made in many firms by reorganising production.
- Government policy: costs will be lowered by government subsidies and raised by various taxes.

The profitability of alternative products (substitutes in supply). If some alternative product (a **substitute in supply**) becomes more profitable to supply than before, producers are likely to switch from the first good to this alternative. Supply of the first good falls. Other goods are likely to become more profitable if their prices rise or their costs of production fall. For example, during 2020 many gin distilleries switched to producing hand sanitiser. Supply of gin fell, while supply of hand sanitiser rose.

The profitability of goods in joint supply. Sometimes when one good is produced, another good is also produced at the same time. These are said to be **goods in joint supply**. An example is the refining of crude oil to produce petrol. Other grade fuels will be produced as well, such as diesel and paraffin. If more petrol is produced, due to a rise in demand, then the supply of these other fuels will rise too.

Nature, 'random shocks' and other unpredictable events. Here we would include the weather and diseases affecting farm output, wars affecting the supply of imported raw materials (such as oil and gas following the Russian invasion of Ukraine), the breakdown of machinery, industrial disputes, floods and of course pandemics, such as COVID-19.

The aims of producers. A profit-maximising firm will supply a different quantity from a firm that has a different aim, such as maximising sales. For most of the time we shall assume that firms are profit maximisers.

Expectations of future price changes. If price is expected to rise, producers may temporarily reduce the amount they sell. Instead they are likely to build up their stocks and release them on to the market only when the price does rise. At the same time, they may plan to produce more, by installing new machines, or taking on more labour, so that they can be ready to supply more when the price has risen.

The number of suppliers. If new firms enter the market, supply is likely to rise.

Pause for thought

By reference to each of the above determinants of supply, identify what would cause (a) the supply of potatoes to fall and (b) the supply of leather to rise.

Definitions

Substitutes in supply These are two goods where an increased production of one means diverting resources away from producing the other.

Goods in joint supply These are two goods where the production of more of one leads to the production of more of the other.

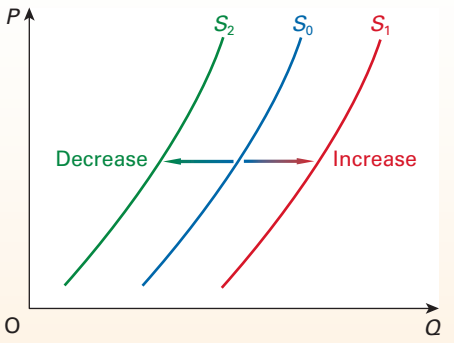
Movements along and shifts in the supply curve

The principle here is the same as with demand curves. The effect of a change in price is illustrated by a movement along the supply curve: for example, from point *d* to point *e* in Figure 2.4 when price rises from 80p to 100p. Quantity supplied rises from 530 000 to 700 000 tonnes per month.

If any other determinant of supply changes, the whole supply curve will shift. A rightward shift illustrates an increase in supply. A leftward shift illustrates a decrease in supply. Thus in Figure 2.4, if the original curve is S_0, the curve S_1 represents an increase in supply (more is supplied at each price), whereas the curve S_2 represents a decrease in supply (less is supplied at each price).

A movement along a supply curve is often referred to as a **change in the quantity supplied**, whereas a shift in the supply curve is simply referred to as a **change in supply**.

Figure 2.4 Shifts in the supply curve



Recap

1. When the price of a good rises, the quantity supplied per period of time will usually also rise. This applies both to individual producers' supply and to the whole market supply.
2. There are two reasons in the short run why a higher price encourages producers to supply more: (a) they are now willing to incur higher costs per unit associated with producing more; (b) they will switch to producing this product instead of now less profitable ones. In the long run there is a third reason: new producers will be attracted into the market.
3. The relationship between price and quantity supplied per period of time can be shown in a table (or schedule) or as a graph. As with a demand curve, price is plotted on the vertical axis and quantity per period of time on the horizontal axis. The resulting supply curve is upward sloping (positively sloped).
4. Other determinants of supply include the costs of production, the profitability of alternative products, the profitability of goods in joint supply, random shocks and expectations of future price changes.
5. If price changes, the effect is shown by a movement along the supply curve. We call this effect 'a change in the quantity supplied'.
6. If any determinant *other* than price changes, the effect is shown by a shift in the whole supply curve. We call this effect 'a change in supply'. A rightward shift represents an increase in supply; a leftward shift represents a decrease in supply.

2.3 THE DETERMINATION OF PRICE

How much of any item will actually be bought and sold and at what price?

Equilibrium price and output

TC 4 p 13

We can now combine our analysis of demand and supply. This will show how the actual price of a product and the actual quantity bought and sold are determined in a free and competitive market.

Let us return to the example of the market demand and market supply of potatoes, and use the data from Tables 2.1 and 2.2. These figures are given again in Table 2.3.

What will be the price and output that actually prevail? If the price started at 20p per kilo, demand would exceed supply by 600 000 tonnes ($A - a$). Consumers would be unable to obtain all they wanted and would thus be willing to pay a higher price. Producers, unable or unwilling to supply enough to meet the demand,

Definitions

Change in the quantity supplied The term used for a movement along the supply curve to a new point. It occurs when there is a change in price.

Change in supply The term used for a shift in the supply curve. It occurs when a determinant *other* than price changes.

Table 2.3 The market demand and supply of potatoes (monthly)

Price of potatoes (pence per kilo)	Total market demand (tonnes: 000s)	Total market supply (tonnes: 000s)
20	700 (*A*)	100 (*a*)
40	500 (*B*)	200 (*b*)
60	350 (*C*)	350 (*c*)
80	200 (*D*)	530 (*d*)
100	100 (*E*)	700 (*e*)

will be only too happy to accept a higher price. The effect of the shortage, then, will be to drive up the price. The same would happen at a price of 40p per kilo. There would still be a shortage; the price would still rise. But as the price rises, the quantity demanded falls and the quantity supplied rises. The shortage is progressively eliminated.

What would happen if the price started at a much higher level: say at 100p per kilo? In this case supply would exceed demand by 600 000 tonnes ($e - E$). The effect of this surplus would be to drive the price down as farmers competed against each other to sell their excess supplies. The same would happen at a price of 80p per kilo. There would still be a surplus; the price would still fall.

In fact, only one price is sustainable. This is the price where demand equals supply: namely 60p per kilo, where both demand and supply are 350 000 tonnes. When supply matches demand the market is said to **clear**. There is no shortage and no surplus.

KEY IDEA 10 TC 8

Equilibrium is the point where conflicting interests are balanced. **Only at this point is the amount that demanders are willing to purchase the same as the amount that suppliers are willing to supply. It is a point which will be automatically reached in a free market through the operation of the price mechanism. This is another of our threshold concepts.**

As we saw on page 23, the price where demand equals supply is called the *equilibrium price*. In Table 2.3, if the price starts at other than 60p per kilo there will be a tendency for it to move towards 60p. The equilibrium price is the only price at which producers' and consumers' wishes are mutually reconciled: where the producers' plans to supply exactly match the consumers' plans to buy.

Demand and supply curves

The determination of equilibrium price and output can be shown using demand and supply curves. Equilibrium is where the two curves intersect. TC 4 p 13

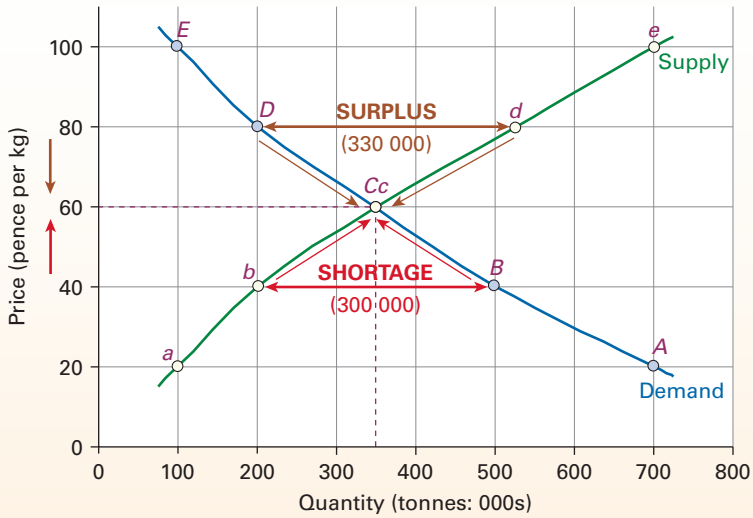
Figure 2.5 shows the demand and supply curves of potatoes corresponding to the data in Table 2.3. Equilibrium price is P_e (60p) and equilibrium quantity is Q_e (350 000 tonnes).

At any price above 60p, there would be a surplus. Thus at 80p there is a surplus of 330 000 tonnes ($d - D$). More is supplied than consumers are willing and able to purchase at that price. Thus a price of 80p fails to clear

Definition

Market clearing A market clears when supply matches demand, leaving no shortage or surplus.

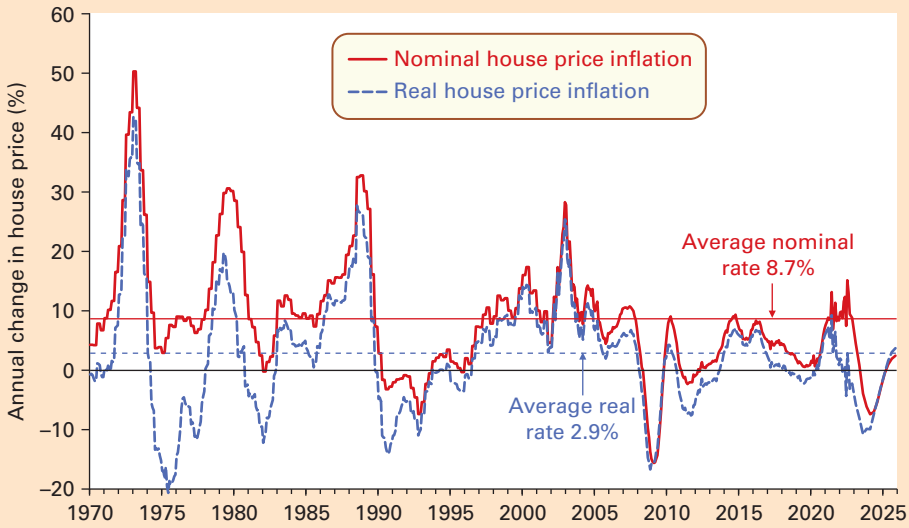
Figure 2.5 The determination of market equilibrium (potatoes: monthly)



BOX 2.1 UK HOUSE PRICES

Unearthing the foundations of house price patterns

UK house price inflation



Sources: Based on data from *Consumer Price Inflation time series dataset* and *UK House Price Index: reports* (Office for National Statistics); forecasts from 2022 Q4 based on data from *Economic and Fiscal Outlook* (Office for Budget Responsibility, November 2022)

In Box 1.3 we saw that the actual (nominal) average UK house price had risen by 6900 per cent between 1970 and 2022. After eliminating general consumer price inflation this represented a 290 per cent *real* increase.

House prices are also volatile. This is readily seen if we plot the annual rates of house price inflation: i.e. the percentage change in house prices over a 12-month period. The chart shows cycles in both nominal and real house prices. Therefore, house prices are not simply volatile because of patterns in general prices; instead they exhibit volatility over and above that in general prices.

The chart shows marked increases in nominal house prices in the early 1970s, with annual house price inflation rates reaching 50 per cent, and again in the early 1980s when rates reached 30 per cent. On both occasions, house price inflation rates soon moderated which, although the actual level of house prices continued to increase, resulted in real house prices falling. The late 1980s saw further very strong growth in house prices in the late 1980s, but this time it was followed by a period of protracted nominal and real house price falls.

By the end of the 1990s house price inflation picked up again and the UK then experienced a protracted period of robust house price growth. But from 2008, as the UK economy entered recession, UK house prices declined again, falling by 16 per cent in the 12 months to February 2009. But then, after an initial surge, house prices remained relatively flat until late 2013, when they began to rise again.

Following the vote in 2016 to leave the EU, house price growth slowed. Then, with the COVID-19 pandemic, house prices fell between March and April 2020 following the first national lockdown, as people were advised not to move house. When lockdown restrictions eased in the second half of 2020, prices rose rapidly, with house price inflation reaching 10.6 per cent in February 2022.

House price volatility

KI 8 p 24

House prices are determined by demand and supply. If demand rises (i.e. shifts to the right) or if supply falls (i.e. shifts to the left), the equilibrium price of houses will rise. Similarly, if demand falls or supply rises, the equilibrium price will fall.

So why have we seen such volatility in house prices? The answer lies largely in fluctuations in *demand*. The following are important factors affecting the demand for houses.

Incomes (actual and anticipated). A growth in real household incomes can encourage people to spend more on housing by taking on larger mortgages, especially if they expect their incomes to continue growing. Examples include the second half of the 1980s and the late 1990s up to 2007, when economic growth was generally strong.

The early 1990s and late 2000s/early 2010s, by contrast, were periods of recession or low economic growth, with rising unemployment and flat or falling real incomes. Consequently, people had much less confidence about their ability to afford large mortgages. In 2017, rising inflation rates again caused real incomes to fall and, with uncertainty about Brexit, house price inflation slowed until 2019.

With COVID-19 causing falling incomes and rising unemployment, people expected house prices to fall during 2020. However, with government measures to support employment, household disposable income fell by just 3.3 per cent. As household spending on commuting, travel and eating out fell, this freed up income that could be spent on housing. Along with other factors outlined below, this allowed house prices to continue rising.

The number of households. Social and demographic changes have resulted in a sharp increase in the number of households over recent decades and hence an increase in demand for housing. In 1981 there were 20 million households in the UK;

by 2021 this had increased to 28.1 million. Reasons include more lone parents, a reduction in average household size, increased life expectancy and inflows of workers from abroad.

The cost of mortgages. During the second half of the 1980s, mortgage interest rates were generally falling, making house purchase more affordable. In 1989, however, this trend went into reverse. Mortgage interest rates started rising. Many people found it difficult to maintain their existing payments, let alone take on a larger mortgage. From 1996 to 2003 mortgage rates generally fell, once more fuelling the demand for houses. Even with gently rising interest rates from 2003 to 2007, mortgages were still relatively affordable. Between 2009 and 2021, interest rates remained at an all-time low because of the uncertainty caused by the slow recovery from the financial crisis, the impact of the EU referendum and then COVID-19. However, interest rates began increasing at the start of 2022 to deal with rising inflation and continued to do so. With a deteriorating macroeconomic environment, house prices were forecast to fall in 2023 and 2024 (see chart).

The availability of mortgages. In the two housing boom periods of the late 1980s and from 1997 to 2007, mortgages were readily available. With house prices rising, banks and building societies were prepared to accept smaller deposits on houses, and to lend a larger multiple of people's income.

In the early 1990s, however, and again from the late 2000s, banks and building societies became more cautious about granting mortgages. They were aware that with falling house prices, rising unemployment and the growing problem of negative equity (people owing more than their house was worth), there was an increased danger that borrowers would default on payments. The problem in the late 2000s was compounded by the financial crisis, which meant that banks had less money to lend. Credit criteria remained tight into the early 2010s, with buyers needing historically large deposits. This significantly reduced the number of first-time buyers. However, the government-backed 'Help to Buy' scheme helped borrowers get a mortgage with a 5 per cent deposit and this contributed to house prices rising once again.

COVID-19 affected lenders in a similar way to earlier recessions, with greater deposits required once more. In response, the government introduced the Mortgage Guarantee Scheme in April 2021, whereby the government would compensate lenders for a proportion of losses in the event of a repossession.

Speculation. A belief that house prices will continue to move in a particular direction can exacerbate house price movements. In other words, speculation tends to increase house price volatility. In the late 1980s and during much of the 2000s there was a belief that house prices would continue to rise. This encouraged people to buy as soon as possible before prices went up any further. There was also an effect on supply. Those with houses to sell held back in the hope of getting a higher price. The net effect was a rightward shift in the demand curve for houses and a leftward shift in the supply curve. The effect of this speculation, therefore, was to help bring about the very effect that people were predicting (see Section 3.5).

In the early 1990s and late 2000s, the opposite occurred. With house prices falling, potential buyers held back, hoping to buy at a lower price. People with houses to sell tried to sell as quickly as possible before prices fell any further. Again, the effect of this speculation was to aggravate the change in prices – this time a fall in prices.

The impact of speculation has also been compounded by the growth in the 'buy-to-let' industry, with mortgage lenders entering this market in large numbers. There was a huge amount of media attention on the possibilities for individuals to make very high returns.

Taxation. In England and Northern Ireland, buyers of property whose price is above a certain threshold have to pay Stamp Duty – a tax on house purchase. During the COVID-19 pandemic, the government increased the threshold from £125 000 to £500 000 in July 2020, which seems to have been a key factor behind the increase in prices in the second half of 2020. The threshold returned to £125 000 in October 2021 (after having been reduced to £250 000 in July 2021).

Long-term house price increases

So why have UK house prices increased both in nominal and real terms over the long term? The answer is that demand has grown faster than supply.

A key factor affecting supply has been the amount of new build: i.e. the volume of residential property construction. Annual levels of new build have generally trended downwards over the past forty to fifty years. In the 1970s, the average level of new build was around 200 000 per year. In the 2010s the figure had fallen around 150 000. As levels of new build have generally fallen, this has put upward pressure on house prices.

Governments' housing policy in recent years has aimed to encourage the building industry by providing tax and other incentives and streamlining planning regulations. But house building may bring adverse environmental and social problems and people often oppose new developments in their area. Increasing pressure for new build on the 'Green Belt' is a particularly controversial issue.

1. *Draw supply and demand diagrams to illustrate what happened to UK house prices (a) in the second half of the 1980s and the period from 1997 to 2007; (b) in the early 1990s and the period from 2008; (c) in the pandemic from 2020 to 2022.*
2. *What determines the supply of housing? How will factors on the supply side influence house prices?*
3. *What is the role of the prices of 'other goods' in determining the demand for housing?*

Undertake an Internet search to find out what forecasters are predicting for house prices over the next year and attempt to explain the role played by demand and supply in their forecasts.

BOX 2.2 STOCK MARKET PRICES

Taking stock of share prices

Firms that are quoted on the stock market (see Case Study 4.1 in MyLab Economics) can raise money by issuing shares. These are sold on the 'primary stock market'. People who own the shares receive a 'dividend' on them, normally paid six-monthly. This varies with the profitability of the company.

People or institutions that buy these shares, however, may not wish to hold on to them for ever. This is where the 'secondary stock market' comes in. It is where existing shares are bought and sold. There are stock markets, primary and secondary, in all the major countries of the world.

At the end of February 2022, there were 2009 companies whose shares are listed on the London Stock Exchange. Shares are traded each Monday to Friday (excluding bank holidays).

KI 8 p 24

The prices of shares depend on demand and supply. For example, if the demand for Tesco shares at any one time exceeds the supply on offer, the price will rise until demand and supply are equal. Share prices fluctuate throughout the trading day and sometimes price changes can be substantial.

To give an overall impression of share price movements, stock exchanges publish share price indices. The best-known example in the UK is the FTSE 100, which stands for the 'Financial Times Stock Exchange' index of the 100 largest companies' shares. The index was first calculated on 3 January 1984 with a base level of 1000 points, and the chart shows the closing monthly value since that time. The chart also shows the Retail Price Index (RPI), re-based so that it too was equal to 1000 in January 1984. The RPI shows weighted average prices of all goods and services. (See Web Appendix 1.1 on the student website for details on the construction and use of price indices.)

From its start in 1984 to early 2022, the FTSE 100 increased by an average of 6.3 per cent per year. However, this figure masks some significant variations in the prices of shares. The index reached a peak of 6930 points on 30 December 1999 and fell to 3287 on 12 March 2003; it then rose again, reaching a high of 6730 on 12 October 2007. Then, with the financial crisis, the index fell to a low of 3512 in March 2009 and an upward trend ensued. The Brexit vote caused a dip, but it quickly recovered, reaching its highest ever closing level of 7877 on 22 May 2018. The rise reflected a general buoyancy in stock markets around the world.

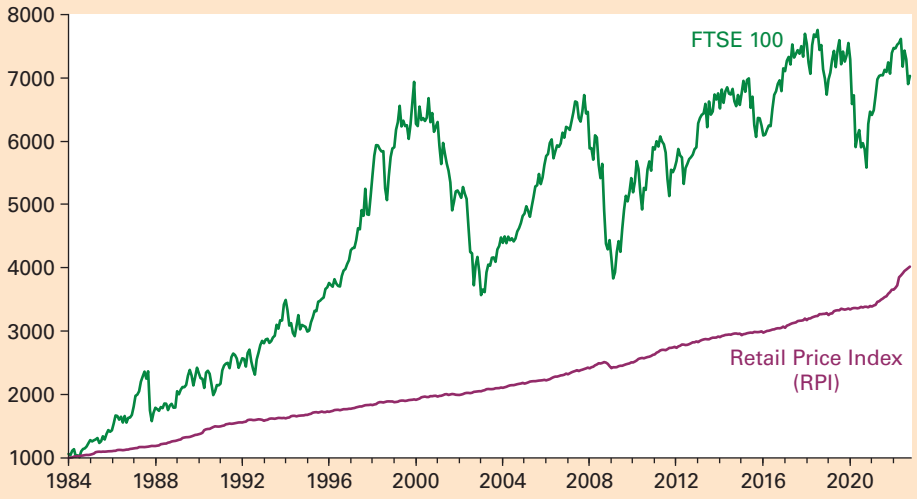
In early 2020, worries about the impact of the COVID-19 pandemic then began to have an impact on investor sentiment. The FTSE 100 fell by 32 per cent in just over a month and dropped below 5000 on 23 March. But then, with the ending of the first lockdown, it climbed to over 6000 in May 2020. The development and successful rollout of the vaccine programme positively affected investor confidence and the index climbed back over 7000 in April 2021 and continued its upward trajectory until early 2022. But then, with fears of rising inflation and interest rates, and expectations over the likely impact of sanctions against Russia over Ukraine, share prices fell rapidly.

When you compare the FTSE 100 with the RPI, you can see that share prices have risen faster than prices generally (the RPI rose by an annual average of 3.6 per cent from 1984 to 2022). Hence, *real* share prices have risen. Also share prices have fluctuated much more than the general level of prices.

The long-run upward trend in real share prices, coupled with their volatility, is something we also see in UK house prices (see Box 2.1).

But what causes share prices to rise and fall? The answer lies in the determinants of the demand and supply of shares.

Financial Times Stock Exchange Index (FTSE 100) (3/1/1984 = 1000)



Note: FTSE figures based on end-of-month values

Sources: Based on data from *Consumer Price Inflation time series dataset*, series *CHAW* (ONS) and various (2022)

Demand

There are five main factors that affect the demand for shares.

The dividend yield. This is the dividend on a share as a percentage of its price. The higher the dividend yields on shares, the more attractive they are as a form of saving. One of the main explanations of rising stock market prices from 2003 to 2007 was high profits and resulting high dividends. By contrast, the slowdowns in the world economy after 2007 and in early 2020 led to falling profits, falling dividends and falling share prices, while the subsequent recoveries caused them to increase once again.

The price of and/or return on substitutes. The main substitutes for shares in specific companies are other shares. Thus if, in comparison with other shares, Tesco shares are expected to pay high dividends relative to the share price, people will buy Tesco shares. For shares in general, the main substitutes are other forms of saving. Thus, if the interest rate on savings accounts in banks and building societies fell, people with such accounts would be tempted to take their money out and buy shares instead.

Another major substitute is property. If house prices rise rapidly, as they did from 1997–2007 and in late 2020 (see Box 2.1), this will reduce the demand for shares as many people switch to buying property in anticipation of even higher prices. If house price growth weakens, as it did in 2018–19, it makes shares relatively more attractive as an investment and can boost their demand.

Of course, other factors may affect *both* house prices *and* shares in the same way. Thus the 'credit crunch' of 2007/8, when finance became both harder and more expensive to obtain, resulted in both falling house prices *and* falling share prices as the economy slowed down and profits fell. Instead, investors looked towards other, safer, investments such as gold, government debt (Treasury bills and gilts) or even holding cash.

But then with interest rates, including those on savings accounts, being dramatically cut as a result of Bank of England measures to stimulate the economy in 2009, many people saw shares as an attractive alternative to bank and building society accounts. The stock market began rising again.

Incomes. If the economy is growing rapidly and people's incomes are thus rising rapidly, they are likely to buy more shares. Thus from 2003–7, when average UK incomes were rising, share prices rose rapidly (see chart). When real incomes fell following the financial crisis in 2008 and again in 2020 with the COVID-19 pandemic, so did share prices.

Wealth. 'Wealth' is people's accumulated savings and property. Wealth rose in the 1990s and 2000s, and many people used their increased wealth to buy shares. The growth in wealth was halted in the late 2000s by the financial crisis and many people looked to 'cash in' their shares, which depressed share prices.

Expectations. In both the mid-to-late 1980s and 1990s, and again from 2003 to 2007, people expected share prices to go on rising. They were optimistic about an end to 'boom and bust' and continued economic growth. These positive sentiments encouraged people to buy shares and pushed share prices up even more, which then further fuelled speculation and further share buying.

In the early 2000s, by contrast, confidence was shaken as growth began to slow. This was exacerbated by other factors, including the 11 September 2001 attack on the World Trade Center and a range of corporate scandals. As share prices fell, so people held back from buying, thereby pushing prices even lower.

A similar thing occurred with the global banking crisis and again with the COVID-19 pandemic. As investor confidence fell and the economy slowed, there was a dramatic fall in share prices. Anticipating further price falls, people held back from buying, thus reducing demand and pushing prices even lower. The success of the vaccine programme in the UK made investors increasingly confident about the chances of an economic recovery, causing share prices to rise consistently once more. However, growing worries about inflation, higher interest rates and the impact of the war in Ukraine caused expectations to become gloomier in early 2022.

The rise and fall in share prices associated with expectations mirror those seen in the housing market and discussed in Box 2.1.

Supply

The factors affecting supply are largely the same as those affecting demand, but in the opposite direction.

If the return on alternative forms of saving falls, people with shares are likely to hold on to them, as they represent a better form of saving. The supply of shares to the market will fall. Similarly, if incomes or wealth rise, people are likely to want to hold on to their shares.

As far as expectations are concerned, if people believe that share prices will rise, they will hold on to the shares they have. Supply to the market will fall, thereby pushing up prices. If, however, they believe that prices will fall, they will sell their shares now before prices do fall. Supply will increase, driving down the price. This happened in 2008, between February and March 2020 and in early 2022. Supply increased and this led to a dramatic fall in the FTSE 100 index.

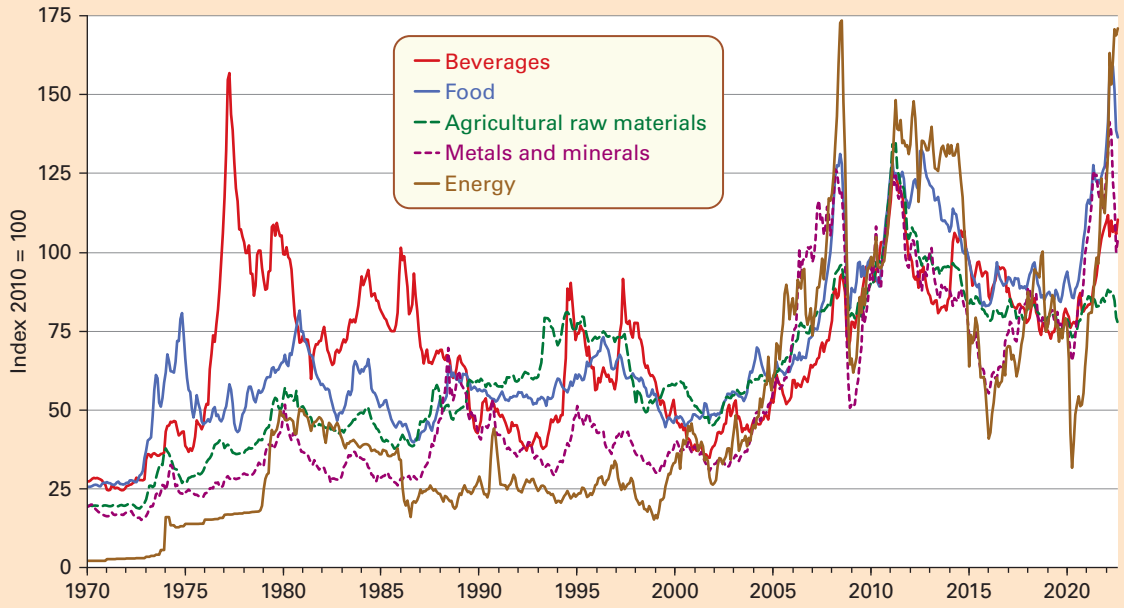
1. *If the rate of economic growth in the economy is 3 per cent in a particular year, can we expect share prices to rise by 3 per cent that year? Explain your answer.*
2. *Why would you expect the return on shares to be greater than that offered by a banks savings account?*

Research what has happened to the FTSE 100 index over the past 12 months (see site B27 on the hotlinks part of the website). Summarise the patterns you find and examine the economic factors helping to drive these patterns.

BOX 2.3 COMMODITY PRICES

Riding the commodities Big Dipper

World primary commodity prices (2010 = 100)



Source: Based on data from *World Bank Commodity Price Data (The Pink Sheet)* (Commodity Markets, World Bank, 2022)

The term 'commodity' in the context of commodity prices refers to the prices of *energy* (e.g. coal, crude oil and natural gas), *foodstuffs* (e.g. cereals, fruits, meat and cooking oils), *beverages* (e.g. coffee, tea and cocoa), *agricultural raw materials* (e.g. timber, cotton, rubber, and tobacco), *metals and minerals* (e.g. copper, aluminium and iron ore) and *fertilisers* (e.g. phosphates and potassium).

The chart shows the pattern of prices of the first five of these since 1970. As you can see in the chart, commodity prices rose strongly in the early 1970s, mid-2000s, the early 2010s and from 2021. There were other periods when prices fell back. For example, commodity prices generally eased between 2012 and 2015.

But, what explains the volatility in commodity prices? To understand this, we need to consider some of the factors that lead to shifts in the demand and supply of commodities. The bigger these shifts, the bigger the resulting changes in commodity prices.

KI 8 p 24

Changes in demand

Income

Economic growth, through its impact on incomes, is a crucial determinant of the demand for certain commodities. The demand for commodities is acutely sensitive to changes in incomes. In recent years, the rate of economic growth in developing countries has been especially important in affecting the demand for commodities.

The rapid rate of growth of developing countries during the 2000s, including the nations collectively known as the BRICS (Brazil, Russia, India, China and South Africa), helped to fuel the demand for raw materials.

With the global economic downturn in 2009, however, particularly in advanced economies, the demand for commodities weakened, which caused commodity prices to decline. While a large number of developing countries continued to experience growth throughout the global downturn, even here growth rates were generally weaker.

Economic growth resumed in many advanced economies in the early 2010s. However, economic growth was to remain fragile for some time. Business and consumer confidence remained weak, while various sectors of the economy, including governments, curbed their borrowing in an attempt to improve their financial positions. These ingredients helped to supress aggregate demand and the general demand for commodities. In due course, this began to spill over to economic growth in developing countries by weakening income flows from export growth, particularly so in economies dependent on earnings from commodities.

the market. Price will fall to the equilibrium price of 60p. As it does so, there will be a movement along the demand curve from point D to point C, and a movement along the supply curve from point d to point c.

At any price below 60p, there would be a shortage. Thus at 40p there is a shortage of 300 000 tonnes ($B - b$). Price will rise to 60p. This will cause a movement along the supply curve from point b to

CASE STUDIES & APPLICATIONS

There was also a weakening in commodity prices at the start of the pandemic in 2020 as the demand for raw materials, and especially energy, fell. But then, as economies recovered, so the demand for commodities increased and prices rose.

Tastes

Changing tastes and trends can also affect prices. The growth of the 'coffee-shop culture' and the fashion for speciality coffees has moved beyond advanced economies and into developing countries. Countries such as China, India and Brazil have seen large increases in consumption. The demand for coffee makes it the second most traded commodity after oil.

Changes in supply

Natural events

The supply of commodities is susceptible to natural events, ranging from droughts or floods to earthquakes. For example, coffee prices in early 2014 soared as Brazil, which produces approximately a third of the world's coffee, saw its worst drought for a decade, and a coffee fungus spread through Central America. Then in 2017/18, the Brazilian harvest was affected by a berry borer beetle infestation. The use of the insecticide (endosulfan) that was traditionally used by growers to control the spread of these beetles was prohibited in 2013 because of its negative impact on human health. (See Case Study 2.5 on the student website for more analysis on coffee market prices.)

Alternative land uses

In some cases, the land used to provide particular commodities has alternative uses. Increases in energy prices, concerns over greenhouse gas emissions and incentives from governments have seen the conversion of agricultural land to the production of biofuels. Ethanol, for instance, can be produced by fermenting the sugars from crops such as wheat and sugar canes and can then be used to power petrol engines. (See Case Study 2.6 on the student website for an examination of the demand for ethanol in Brazil.)

Changes in the sources of supply

The fall in the price of energy (fuels) from summer 2014 was partly the result of faltering global demand but mainly the result of increased supply from shale oil deposits, particularly in the USA. The increased supply of shale oil was not offset by a reduction in other production. From over $100 per barrel at the start of 2014, the price of oil dipped below $30 per barrel in early 2016. This also reflected the then strategy of OPEC (the Organisation of the Petroleum Exporting Countries) to maintain lower prices to make shale oil production increasingly less profitable to discourage further shale exploration and investment. However, the effect on export earnings was not sustainable. In December 2016, OPEC and Russia agreed to cut production.

A large reduction in effective supply occurred in 2022 with the Russian invasion of Ukraine. Western countries, as part of their sanctions, announced that they would phase out the use of Russian oil and gas. This effective reduction in the sources of supply led to a large rise in energy prices. This compounded the temporary reductions in supply that had occurred as a result of supply-chain problems in the wake of the COVID-19 pandemic.

Speculation (affecting both demand and supply)

Many financial institutions, companies and individuals speculate in commodities, hoping to make money by buying at a low price and selling at a high price. When successful, speculators can make large percentage gains in a short period of time. However, they can also lose by getting their predictions wrong.

In uncertain times, speculation can be destabilising, exaggerating price rises and falls as speculators 'jump on the bandwagon', seeing price changes as signifying a trend. In more stable times, speculation can even reduce price fluctuations as speculators buy when prices are temporarily low and sell when they are temporarily high. We examine speculation in more detail in Section 3.5.

During particularly uncertain times, such as those in the late 2000s, early 2010s and early 2020s, confidence and, hence, commodity prices, fluctuated significantly. On days of good economic news demand for commodities would tend to rise as people believed that a growing world economy would drive up the demand for commodities and hence their prices. On days of bad economic news, the price of commodities would tend to fall as the demand for commodities fell. Similarly, serious worries about disruptions to supply, as occurred in early 2022 with first the supply-chain problems and then the Russian invasion of Ukraine, can drive up commodity prices significantly.

1. *What might explain why the price of some commodity types rose less than others during the period of strong global economic growth between 2000 and 2007?*
2. *Explain how speculation can contribute to the volatility of commodity prices.*
3. *Would you expect commodity prices to fluctuate more than house prices?*

Extract economic growth data (growth in GDP at constant prices) from the most recent IMF World Economic Outlook Database for the nations collectively known as the BRICS and for the world economy (see also site B31 on the hotlinks part of the website). Compare the patterns in economic growth since 2000 with those in commodity prices (see the chart in this box) and summarise your findings.

point *c* and along the demand curve from point *B* to point *C*.

Point *Cc* is the equilibrium: where demand equals supply.

Movement to a new equilibrium

The equilibrium price will remain unchanged only so long as the demand and supply curves remain unchanged.

KI 8 p 24

Figure 2.6 Effect of a shift in the demand curve

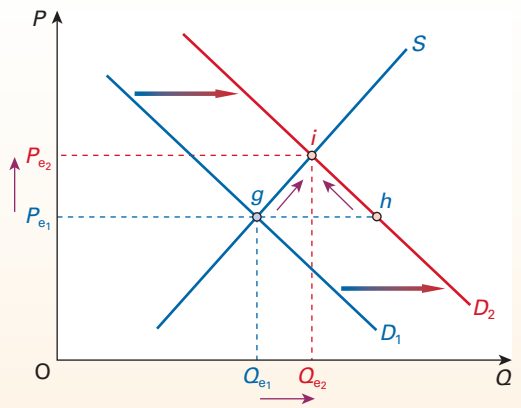
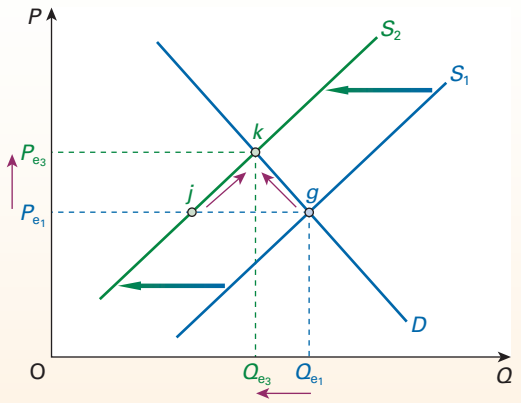


Figure 2.7 Effect of a shift in the supply curve



If either of the curves shifts, a new equilibrium will be formed.

A change in demand

If one of the determinants of demand changes (other than price), the whole demand curve will shift. This will lead to a movement along the supply curve to the new intersection point.

For example, in Figure 2.6, if a rise in consumer incomes led to the demand curve shifting to D_2, there would be a shortage of $h-g$ at the original price P_{e1}. This would cause price to rise to the new equilibrium P_{e2}. As it did so, there would be a movement along the supply curve from point g to point i, and along the new demand curve (D_2) from point h to point i. Equilibrium quantity would rise from Q_{e1} to Q_{e2}.

The effect of the shift in demand, therefore, has been a movement along the supply curve from the old equilibrium to the new: from point g to point i.

A change in supply

Likewise, if one of the determinants of supply changes (other than price), the whole supply curve will shift. This will lead to a movement *along* the *demand* curve to the new intersection point.

For example, in Figure 2.7, if costs of production rose, the supply curve would shift to the left: to S_2. There would be a shortage of $g-j$ at the old price of P_{e1}. Price would rise from P_{e1} to P_{e3}. Quantity would fall from Q_{e1} to Q_{e3}. In other words, there would be a movement along the demand curve from point g to point k, and along the new supply curve (S_2) from point j to point k.

Pause for thought

Is the following statement true? 'An increase in demand will cause an increase in price. This increase in price will cause a reduction in demand, until demand is reduced back to its original level.' Explain your answer and try using a demand and supply diagram to illustrate what is going on.

To summarise: a shift in one curve leads to a movement along the other curve to the new intersection point. Sometimes a number of determinants might change. This may lead to a shift in *both* curves. When this happens, equilibrium simply moves from the point where the old curves intersected to the point where the new ones intersect.

Recap

1. If the demand for a good exceeds the supply, there will be a shortage. This will lead to a rise in the price of the good.
2. If the supply of a good exceeds the demand, there will be a surplus. This will lead to a fall in the price.
3. Price will settle at the equilibrium. The equilibrium price is the one that clears the market: the price where demand equals supply.
4. If the demand or supply curve shifts, this will lead either to a shortage or to a surplus. Price will therefore either rise or fall until a new equilibrium is reached at the position where the supply and demand curves now intersect.

2.4 THE FREE-MARKET ECONOMY

How well does it serve us?

Advantages of a free-market economy

The fact that a free-market economy functions automatically is one of its major advantages. There is no need for costly and complex bureaucracies to co-ordinate economic decisions. The economy can respond quickly to changing demand and supply conditions.

When markets are highly competitive, no one has great power. Competition between firms keeps prices down and acts as an incentive to firms to become more efficient. The more firms there are competing, the more responsive they will be to consumer wishes.

The more efficiently firms can combine their factors of production, the more profit they will make. The more efficiently workers work, the more secure will be their jobs and the higher their wages. The more carefully consumers decide what to buy, the greater the value for money they will receive.

TC 8 p 37

Thus people pursuing their own self-interest through buying and selling in competitive markets helps to minimise the central economic problem of scarcity, by encouraging the efficient use of the nation's resources in line with consumer wishes. From this type of argument, the following conclusion is often drawn by defenders of the free market: 'The pursuit of private gain results in the social good.' This is obviously a highly significant claim and has profound moral implications (see Threshold Concepts 8 and 9).

KEY IDEA 11

TC 9

People gain from voluntary interaction. **When people buy from or sell to other people, or when they are employed by or employ other people, both parties will gain from the interaction. This is the seventh of our threshold concepts.**

Problems with a free-market economy

In practice, however, markets do not achieve maximum efficiency in the allocation of scarce resources, and governments feel it necessary to intervene to rectify this. This is the allocative role for government that we identified in Section 1.5 along with other potential roles for government (see page 26). The problems of a free market include the following:

- Competition between firms is often limited. A few giant firms may dominate an industry. In these cases, they may charge high prices and make large profits. TC 6 p 24
- Rather than merely responding to consumer wishes, firms may attempt to persuade consumers by advertising. Consumers are particularly susceptible to advertisements for products that are unfamiliar to them.
- Lack of competition and high profits may remove the incentive for firms to be efficient.
- Power and property may be unequally distributed. Those who have power and/or property (e.g. big business, unions, landlords) will gain at the expense of those without power and property.
- The practices of some firms may be socially undesirable or have adverse environmental consequences. For example, a chemical works may pollute the environment.
- Some socially desirable goods would simply not be produced by private enterprise. Who would carry out national counter-terrorism activities if this were not funded by governments?
- A free-market economy may lead to macroeconomic instability. There may be periods of recession with high unemployment and falling output, and other periods of very strong economic growth but with rapidly rising prices.
- Finally, there is the ethical objection that a free-market economy, by rewarding self-interested behaviour, may encourage selfishness, greed, materialism and the acquisition or pursuit of power. TC 6 p 24

We shall be examining these various problems in more detail in later chapters and the ways governments intervene to tackle them.

Pause for thought

Go through the potential problems of a free market listed here and identify which of the roles for government (outlined on page 26) each is consistent with.

Recap

1. A free-market economy functions automatically, and if there is plenty of competition between producers, this can help to protect consumers' interests.
2. In practice, however, competition may be limited: there may be great inequality; there may be adverse social and environmental consequences; there may be macroeconomic instability. Consequently, governments intervene in market economies in various ways in order to correct the failings of the free market.

QUESTIONS

1. Why do the prices of fresh vegetables fall when they are in season? Could an individual farmer prevent the price falling?
2. If you were the owner of a clothes shop, how would you set about deciding what prices to charge for each garment at the end-of-season sale?
3. The number of owners of smartphones has grown rapidly and hence the demand for smartphones has also grown rapidly. Yet the prices of smartphones have fallen. Why?
4. Assume that oil begins to run out and that extraction becomes more expensive. Trace through the effects of this on the market for oil and the market for other fuels.
5. This question is concerned with the supply of oil for central heating. In each case consider whether there is a movement along the supply curve (and in which direction) or a shift in it (and whether left or right).
   a. New oil fields start up in production.
   b. The demand for central heating rises.
   c. The price of gas falls.
   d. Oil companies anticipate an upsurge in demand for central heating oil.
   e. The demand for petrol rises.
   f. New technology decreases the costs of oil refining.
   g. All oil products become more expensive.
6. The weekly demand and supply schedules for t-shirts (in millions) in a free market are as follows:

Price (£)	8	7	6	5	4	3	2	1
Quantity demanded	6	8	10	12	14	16	18	20
Quantity supplied	18	16	14	12	10	8	6	4

   a. What is the equilibrium price and quantity?
   b. Assume that changes in fashion cause the demand for t-shirts to rise by 4 million at each price. What will be the new equilibrium price and quantity? Has equilibrium quantity risen as much as the rise in demand? Explain why or why not.
   c. Now plot the data in the table on a graph and mark the equilibrium. Also plot the new data corresponding to (b) and mark the new equilibrium.
7. On separate demand and supply diagrams for bread, sketch the effects of the following: (a) a rise in the price of wheat; (b) a rise in the price of butter and margarine; (c) a rise in the price of rice, pasta and potatoes. In each case, state your assumptions.
8. For what reasons might the price of foreign holidays rise? In each case identify whether these are reasons affecting demand or supply (or both). What happened to the price of foreign holidays during the pandemic and why?
9. If both demand and supply change, and if we know in which direction they have shifted but not how much, why is it that we will be able to predict the direction in which *either* price *or* quantity will change, but not both? (Clue: consider the four possible combinations and sketch them if necessary: (a) *D* left, *S* left; (b) *D* right, *S* right; (c) *D* left, *S* right; (d) *D* right, *S* left.)
10. What will happen to the equilibrium price and quantity of butter in each of the following cases? You should state whether demand or supply (or both) have shifted and in which direction. (In each case assume *ceteris paribus*.)
    a. A rise in the price of non-dairy spreads.
    b. A rise in the demand for yoghurt.
    c. A rise in the price of bread.
    d. A rise in the demand for bread.
    e. An expected rise in the price of butter in the near future.
    f. A tax on butter production.
    g. The invention of a new, but expensive, process for removing all cholesterol from butter, plus the passing of a law which states that all butter producers must use this process.
11. Assume that fracking becomes common across the UK. The result is that supplies of shale gas and oil increase sharply. Trace through the effects of this on the market for oil, gas and the market for other fuels.

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

2.1 **The interdependence of markets.** A case study in the operation of markets, examining the effects on a local economy of the discovery of a large coal deposit.

2.2 **Stocks and flows.** This examines one of the most important distinctions in economics and one that we shall come across on several occasions.

2.3 **Adjusting to oil price shocks.** A case study showing how demand and supply analysis can be used to examine the price changes in the oil market since 1973.

2.4 **Rationing.** A case study in the use of rationing as an alternative to the price mechanism. In particular, it looks at the use of rationing in the UK during the Second World War.

2.5 **Coffee prices.** An examination of the coffee market and the implications of fluctuations in the coffee harvest for growers and coffee drinkers.

2.6 **Response to changes in petrol and ethanol prices in Brazil.** This case examines how drivers with 'Flex-fuel' cars responded to changes in the relative prices of two fuels: petrol and ethanol (made from cane sugar).

Chapter 3

Markets in action

In this chapter we explore the working of markets in more detail. We start by examining one of the most important concepts in the whole of economics – that of elasticity (Sections 3.1–3.4).

A bumper harvest may seem like good news for farmers: after all, they will be able to sell more. But is it good news? Although they will sell more, the effect of the increased supply will be to drive down the price – and that's bad news for farmers! So will the increased sales (the good news) be enough to compensate for the reduction in price (the bad news)? Will farmers end up earning more or less from their bumper harvest? It all depends on just how much the price falls, and this depends on the *price elasticity of demand* for their produce. This is a measure of how *responsive* demand is to a change in price.

It is not just the responsiveness of *demand* that is important in determining the functioning of markets. It is also the responsiveness of *supply*. Why do some firms respond to a rise in price by producing a lot more, whereas others produce only a little more? Is it simply because of different technologies? We will discover just what influences the price elasticity of supply in Section 3.3.

In Section 3.5 we compare the adjustment paths of markets across time to changes in demand or supply. We see how important elasticities are in affecting the nature of this adjustment as well as the role played by expectations. In some markets expectations can help to fuel volatility, while in others they dampen volatility.

The chapter closes by looking at what happens if governments set about *controlling* prices. Why will shortages occur if the government sets the price too low, or surpluses if it sets it too high? When might governments feel that it is a good idea to fix prices?

After studying this chapter, you should be able to answer the following questions:

- How responsive is consumer demand to changes in prices and changes in incomes?
- How responsive is firms' output to a change in price?
- How does this responsiveness (or 'elasticity') of demand and supply affect the working of markets?
- Why are markets likely to be more responsive in the long run than the short run to changes in demand or supply?
- How will people respond if they anticipate a change in price?
- What will happen if the government sets a price either above or below the market equilibrium?

3.1 PRICE ELASTICITY OF DEMAND

How responsive is demand to a change in price?

Chapter 2 introduced one of the most important theories in economics: the theory of demand and supply. We saw how their interaction determines price.

We also saw how, when the price of a good changes, the quantity demanded will change. A rise in price causes the quantity demanded to fall; a fall in price causes the quantity demanded to rise. But in most cases, we will want to know more than this. We will want to know just *how much* the quantity demanded will change. In other words, we will want to know how *responsive* demand is to a change in price.

KI 8 p 24

A change in price also affects the quantity supplied. A rise in price causes the quantity supplied to rise; a fall in price causes the quantity supplied to fall. But again, in most cases we will want to know just *how much* the quantity supplied will change – how *responsive* supply is to a change in price.

The responsiveness of variables like demand and supply to changes in other variables is known as their **elasticity.** Understanding elasticities and what determines their size helps us to gain a better understanding of how particular markets or different sectors of the economy are likely to respond to the ever-changing circumstances of the real world. Elasticity is therefore our next threshold concept.

KEY IDEA 12

TC 10

Elasticity. **The responsiveness of one variable (e.g. demand) to a change in another (e.g. price). This concept is fundamental to understanding how markets work and is thus one of our threshold concepts (no. 10). The more elastic variables are, the more responsive is the market to changing circumstances.**

As we go through the text, we will see that the concept of elasticity can be applied to relationships across economics, both in microeconomics and in macroeconomics. In this chapter we discuss some of the most well-known elasticities: those relating to demand and supply. We begin with the price elasticity of demand.

The price elasticity of demand

The law of demand describes a negative relationship between price and quantity demanded. But this is a general relationship. It does not convey the degree to which quantity demand responds to a change in price.

Take the case of two products: oil and carrots. In the case of oil, a rise in price is likely to result in only a slight fall in the quantity demanded. If people want to continue driving, they have to pay the higher prices for fuel. A few may turn to riding bikes, and some people may try to make fewer journeys, but for most people, a rise in the price of petrol and diesel will make little difference to how much they use their cars.

In the case of carrots, however, a rise in price may lead to a substantial fall in the quantity demanded. The reason is that there are alternative vegetables that people can buy. Many people, when buying vegetables, are very conscious of their prices and will buy whatever is reasonably priced.

We call the responsiveness of demand to a change in price the **price elasticity of demand**. If we know the price elasticity of demand for a product, we can predict the effect on price and quantity of a shift in the *supply* curve for that product. For example, we can predict the effect of the bumper harvest that we considered at the beginning of the chapter.

TC 4 p 13

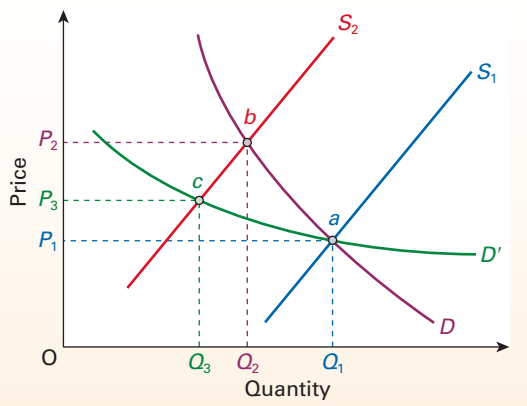
Figure 3.1 shows the effect of a shift in supply with two quite different demand curves (D and D'). Curve D' is more elastic than curve D over any given price range. In other words, for any given change in price, there will be a larger change in quantity demanded along curve D' than along curve D.

Assume that initially the supply curve is S_1, and that it intersects with both demand curves at point a, at a price of P_1 and a quantity of Q_1. Now supply shifts to S_2.

Definition

Price elasticity of demand The responsiveness of quantity demanded to a change in price.

Figure 3.1 Market supply and demand



What will happen to price and quantity? In the case of the less elastic demand curve D, there is a relatively large rise in price (to P_2) and a relatively small fall in quantity (to Q_2): equilibrium is at point b. In the case of the more elastic demand curve D', however, there is only a relatively small rise in price (to P_3) but a relatively large fall in quantity (to Q_3): equilibrium is at point c.

Measuring the price elasticity of demand

What we want to compare is the size of the change in quantity demanded with the size of the change in price. But since price and quantity are measured in different units, the only sensible way we can do this is to use percentage or proportionate changes. This gives us the following **formula for the price elasticity of demand** ($P\epsilon_D$) for a product: percentage (or proportionate) change in quantity demanded divided by the percentage (or proportionate) change in price. Putting this in symbols gives:

$$P\epsilon_D = \frac{\%\Delta Q_D}{\%\Delta P}$$

where ϵ (the Greek epsilon) is the symbol we use for elasticity, and Δ (the capital Greek delta) is the symbol we use for a 'change in'.

Thus if a 50 per cent rise in the price of oil caused the quantity demanded to fall by a mere 10 per cent, the price elasticity of oil over this range would be:

$$-10\%/50\% = -0.2$$

On the other hand, if a 5 per cent fall in the price of carrots caused a 20 per cent rise in the quantity demanded, the price elasticity of demand for carrots over this range would be:

$$20\%/-5\% = -4$$

Carrots have a more elastic demand than oil, and this is shown by the figures. But just what do these two figures show? What is the significance of minus 0.2 and minus 4?

Interpreting the figure for elasticity

The use of proportionate or percentage measures

Elasticity is measured in proportionate or percentage terms for the following reasons:

- It allows comparison of changes in two qualitatively different things, and thus which are measured in two different types of unit: i.e. it allows comparison of *quantity* changes with *monetary* changes.
- It is the only sensible way of deciding *how big* a change in price or quantity is. Take a simple example. An item goes up in price by £1. Is this a big increase or a small increase? We can answer this only if we know what the original price was. If a can of beans goes up in price by £1, that is a huge price increase. If, however, the price of a house goes up by £1, that is a tiny price increase. In other words, it is the percentage or proportionate increase in price that we look at in deciding how big a price rise it is.

The sign (positive or negative)

Demand curves are generally downward sloping. This means that price and quantity change in opposite directions. A *rise* in price (a positive figure) will cause a *fall* in the quantity demanded (a negative figure). Similarly, a *fall* in price will cause a *rise* in the quantity demanded. Thus when working out price elasticity of demand we either divide a negative figure by a positive figure, or a positive figure by a negative. Either way, we end up with a negative figure.

The value (greater or less than 1)

If we now ignore the negative sign and just concentrate on the value of the figure, this tells us whether demand is elastic or inelastic.

- **Elastic demand** ($\epsilon > 1$). This is where a change in price causes a proportionately larger change in the quantity demanded. In this case the value of elasticity will be greater than 1, since we are dividing a larger figure by a smaller figure.
- **Inelastic demand** ($\epsilon < 1$). This is where a change in price causes a proportionately smaller change in the quantity demanded. In this case elasticity will be less than 1, since we are dividing a smaller figure by a larger figure.
- **Unit elastic demand** ($\epsilon = 1$). This is where price and quantity demanded change by the same proportion. This will give an elasticity equal to 1, since we are dividing a figure by itself.

Definitions

Formula for price elasticity of demand ($P\epsilon_d$) The percentage (or proportionate) change in quantity demanded divided by the percentage (or proportionate) change in price: $\%\Delta Q_D \div \%\Delta P$.

Elastic demand Where quantity demanded changes by a larger percentage than price. Ignoring the negative sign, it will have a value greater than 1.

Inelastic demand Where quantity demanded changes by a smaller percentage than price. Ignoring the negative sign, it will have a value less than 1.

Unit elastic demand Where quantity demanded changes by the same percentage as price. Ignoring the negative sign, it will have a value equal to 1.

Determinants of price elasticity of demand

The price elasticity of demand varies enormously from one product to another. But why do some products have a highly elastic demand, whereas others have a highly *in*elastic demand? What determines price elasticity of demand?

The number and closeness of substitute goods. This is the most important determinant. The more substitutes there are for a good, and the closer they are, the more people will switch to these alternatives when the price of the good rises: the greater, therefore, will be the price elasticity of demand.

Pause for thought

Why will the price elasticity of demand for holidays in Spain be greater than that for holidays in general? Is this difference the result of a difference in the size of the income effect or the substitution effect? Is there anything the suppliers of holidays in Spain can do to reduce this higher price elasticity?

A broadly defined market, such as alcoholic drinks, has few substitutes – the obvious one in this case being soft drinks. Customers tend to be relatively insensitive to the price. For example, a number of international studies have found that alcoholic drinks as a whole have a relatively inelastic demand of around –0.4.[1]

A more narrowly defined market, such as beer, is likely to have more substitutes (i.e. cider, wine and spirits) and so consumer demand will tend to be less price inelastic. A figure of between −0.98 and −1.27 has been estimated for off-trade beer: i.e. beer sold in supermarkets and off-licences (bottle shops).[2] The elasticity of demand for a good produced by a single firm (i.e. a particular brand of beer) is likely to be even greater Consumers can switch to another brand of the same product.

The proportion of income spent on the good. The higher the proportion of our income we spend on a good, the more we will be forced to cut consumption when its price rises: the bigger will be the income effect (see page 31) and the more elastic will be the demand.

Thus salt has a very low price elasticity of demand – estimates suggest it is about –0.1.[3] We spend such a tiny fraction of our income on salt that we would find little difficulty in paying a relatively large percentage increase in its price: the income effect of a price rise would be very small. By contrast, there will be a much bigger income effect when the prices of a major item of expenditure rises. For example, if mortgage interest rates rise (the 'price' of loans for house purchase), people may have to cut down substantially on their demand for housing, being forced to buy somewhere smaller and cheaper, or to live in rented accommodation.

The time period. When price rises, people may take time to adjust their consumption patterns and find alternatives. The longer the time period after a price change, then, the more elastic is the demand likely to be.

To illustrate this, let us return to our example of oil. The Office for Budget Responsibility estimates that the price elasticity of demand for road fuel is −0.07 in the short run and −0.13 in the medium term. Other studies have estimated a long-run figure of approximately −0.85.[4]

Why is the figure for fuel so much less elastic in the short run than the long run? If fuel prices rise, as they did in 2022 with the disruption of oil supplies associated with the Russian invasion of Ukraine, people will find it difficult to reduce their consumption by a significant amount in the short run. If public transport options are limited, they still have to drive their cars to work and for leisure purposes. Firms still have to use fuel to transport their goods.

Over time, people can find other ways to respond, such as purchasing electric cars, car sharing or moving closer to their work. There is also an incentive for firms to develop cars that are more fuel-efficient or, in the case of electric cars, have greater battery capacity, or use alternative fuels, such as biofuels or hydrogen. Firms can also change their production methods to economise on oil consumption and find alternative ways to transport their goods.

Pause for thought

Demand for oil might be relatively elastic over the longer term, and yet it could still be observed that over time people consume more oil (or only very slightly less) despite rising oil prices. How can this apparent contradiction be explained?

[1] See, for example: João Sousa, 'Estimate of price elasticities of demand for alcohol in the United Kingdom', *HMRC Working Paper 16*, HM Revenue & Customs (December 2014).

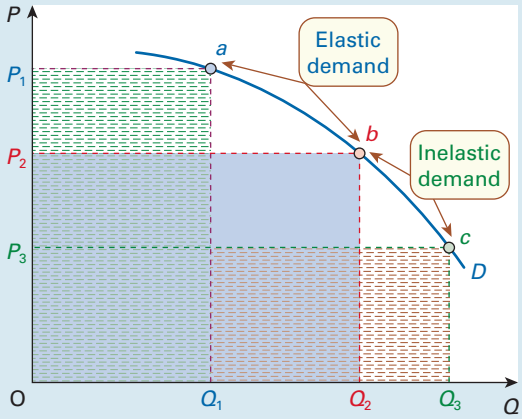
[2] Y. Meng, A. Brennan, R. Purshouse et al. 'Estimation of own and cross price elasticities of alcohol demand in the UK – A pseudo-panel approach using the Living Costs and Food Survey 2001–2009', *Journal of Health Economics*, 34 (2014).

[3] Patrick L. Anderson et al., *Price elasticity of demand*, Mackinac Center for Public Policy (13 November 1997).

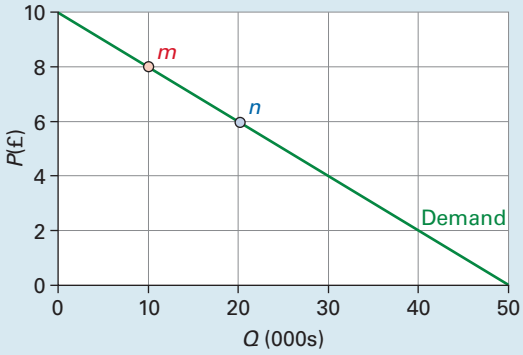
[4] *Analysis of the dynamic effects of fuel duty reductions*, HM Treasury (April 2014).

BOX 3.1 THE MEASUREMENT OF ELASTICITY

(a) Different elasticities along different portions of a demand curve



(b) Measuring elasticity using the arc method



We have defined price elasticity as the percentage or proportionate change in quantity demanded divided by the percentage or proportionate change in price. But how, in practice, do we measure these changes for a specific demand curve?

A common mistake that students make is to think that you can talk about the elasticity of a whole *curve*. The mistake here is that in most cases the elasticity will vary along the length of the curve.

Take the case of the demand curve illustrated in Figure (a). Between points *a* and *b*, total expenditure rises ($P_2Q_2 > P_1Q_1$): demand is thus elastic between these two points. Between points *b* and *c*, however, total revenue falls ($P_3Q_3 < P_2Q_2$). Demand here is inelastic.

Normally, then, we can refer to the elasticity only of a *portion* of the demand curve, not of the *whole* curve.

There are, however, two exceptions to this rule. The first is when the elasticity just so happens to be the same all the

Recap

1. Elasticities capture the responsiveness of variables, such as demand and supply, to changes in other variables, such as price. By analysing elasticities we can gain a better understanding of economic relationships and how economies respond to changing circumstances.
2. Price elasticity of demand is a measure of the responsiveness of demand to a change in price.
3. It is defined as the proportionate (or percentage) change in quantity demanded divided by the proportionate (or percentage) change in price. Given that demand curves are downward sloping, price elasticity of demand will have a negative value.
4. If quantity changes proportionately more than price, the figure for elasticity will be greater than 1 (ignoring the sign): demand is elastic. If the quantity changes proportionately less than price, the figure for elasticity will be less than 1 (again, ignoring the sign): demand is inelastic. If quantity and price change by the same proportion, the elasticity has a value of (minus) 1: demand is unit elastic.
5. Demand will be more elastic the greater the number and closeness of substitute goods, the higher the proportion of income spent on the good and the longer the time period that elapses after the change in price.

3.2 PRICE ELASTICITY OF DEMAND AND CONSUMER EXPENDITURE

How much do we spend on a good at a given price?

One of the most important applications of price elasticity of demand concerns its relationship with the total amount of money that consumers spend on a product. **Total consumer expenditure** *(TE)* is simply price times quantity purchased.

Definition

Total consumer expenditure on a product (*TE*) **(per period of time)** The price of the product multiplied by the quantity purchased: $TE = P \times Q$.

EXPLORING ECONOMICS

way along a curve, as in the three special cases illustrated in Figure 3.3. The second is where two curves are drawn on the same diagram, as in Figure 3.1. Here we can say that demand curve D is less elastic than demand curve D' at any given price. Note, however, that each of these two curves will still have different elasticities along its length.

Although we cannot normally talk about the elasticity of a whole curve, we can nevertheless talk about the elasticity between any two points on it. Remember the formula we used was:

$$\frac{\%\text{or proportionate } \Delta Q}{\%\text{or proportionate } \Delta P} \text{ (where } \Delta \text{ means 'change in')}$$

The way we measure a *proportionate* change in quantity is to divide that change by the level of Q: i.e. $\Delta Q/Q$. Similarly, we measure a proportionate change in price by dividing that change by the level of P: i.e. $\Delta P/Q$. Price elasticity of demand can thus now be rewritten as:

$$\frac{\Delta Q}{Q} \div \frac{\Delta P}{P}$$

But just what value do we give to P and Q? Consider the demand curve in Figure (b). What is the elasticity of demand between points m and n? Price has fallen by £2 (from £8 to £6), but what is the proportionate change? Is it 2/8 or 2/6? The convention is to express the change as a proportion of the average of the two prices, £8 and £6: in other words, to take the mid-point price, £7. Thus the proportionate change is 2/7.

Similarly, the proportionate change in quantity between points m and n is 10/15, since 15 is mid-way between 10 and 20.

Thus using the **average (or 'mid-point') formula**, elasticity between m and n is given by:

$$\frac{\Delta Q}{\text{average } Q} \div \frac{\Delta P}{\text{average } P} = \frac{10}{15} \div \frac{-2}{7} = -2.33$$

Since 2.33 is greater than 1, demand is elastic between m and n.

Referring again to Figure (b), what is the price elasticity of demand between a price of (a) £6 and £4; (b) £4 and £2? What do you conclude about the elasticity of a straight-line demand curve as you move down it?

In this box we have looked at the measurement of elasticity over a segment of the demand curve. This gives what is known as 'arc elasticity'. An alternative is to measure elasticity at a single point on the demand curve. This gives, not surprisingly, what is known as 'point elasticity'. This method is examined in Web Appendix 3.1.

Definition

Average (or 'mid-point') formula for price elasticity of demand $\Delta Q_D/\text{average } Q_D \div \Delta P/\text{average } P$.

$$TE = P \times Q$$

For example, if consumers buy 3 million units (Q) at a price of £2 per unit (P), they will spend a total of £6 million (TE). Note that total consumer expenditure will be the same as the **total revenue** (***TR***) received by firms from the sale of the product (before any taxes or other deductions).

What will happen to consumer expenditure (and hence firms' revenue) if there is a change in price? The answer depends on the price elasticity of demand.

Elastic demand

As price rises so quantity demanded falls, and vice versa. When demand is elastic, quantity demanded changes proportionately more than price. Thus the change in quantity has a bigger effect on total consumer expenditure than does the change in price. For example, when the price rises, there will be such a large fall in consumer demand that less will be spent than before. This can be summarised as follows:

- P rises; Q falls proportionately more: therefore TE falls.
- P falls; Q rises proportionately more: therefore TE rises.

In other words, total expenditure changes in the same direction as *quantity*.

This is illustrated in Figure 3.2(a). The areas of the rectangles in the diagram represent total expenditure. But why? The area of a rectangle is its height multiplied by its length. In this case, this is price multiplied by quantity purchased, which, as we have seen, gives total expenditure.

Demand is elastic between points a and b. A rise in price from £4 to £5 causes a proportionately larger fall in quantity demanded: from 20 million to 10 million. Total expenditure *falls* from £80 million (the striped area) to £50 million (the lilac shaded area)

When demand is elastic, then a rise in price will cause a fall in total consumer expenditure and thus a fall in the total revenue that firms selling the product receive. A reduction in price, however, will result in consumers spending more, and hence firms earning more.

Definition

Total revenue (*TR*) (per period) The total amount received by firms from the sale of a product, before the deduction of taxes or any other costs. The price multiplied by the quantity sold. $TR = P \times Q$.

Pause for thought

If a firm faces an elastic demand curve, why will it not necessarily be in the firm's interests to produce more? (Clue: you will need to distinguish between revenue and profit. We will explore this relationship in Chapter 5.)

Inelastic demand

When demand is inelastic, it is the other way around. Price changes proportionately more than quantity. Thus the change in price has a bigger effect on total consumer expenditure than does the change in quantity. To summarise the effects:

- P rises; Q falls proportionately less: therefore TE rises.
- P falls; Q rises proportionately less: therefore TE falls.

In other words, total consumer expenditure changes in the same direction as *price*.

This is illustrated in Figure 3.2(b). Demand is inelastic between points a and c. A rise in price from £4 to £8 causes a proportionately smaller fall in quantity demanded: from 20 million to 15 million. Total expenditure *rises* from £80 million (the striped area) to £120 million (the lilac shaded area).

In this case, firms' revenue will increase if there is a rise in price, and fall if there is a fall in price.

Pause for thought

Assume that demand for a product is inelastic. Will consumer expenditure go on increasing as price rises? Would there be any limit?

Special cases

Figure 3.3 shows three special cases: (a) a totally inelastic demand ($P\epsilon_D = 0$); (b) an infinitely elastic demand ($P\epsilon_D = -\infty$); and (c) a unit elastic demand ($P\epsilon_D = -1$).

Totally inelastic demand. This is shown by a vertical straight line. No matter what happens to price, quantity demanded remains the same. It is obvious that the more the price rises, the bigger will be the level of consumer expenditure. Thus in Figure 3.3(a) consumer expenditure will be higher at P_2 than at P_1.

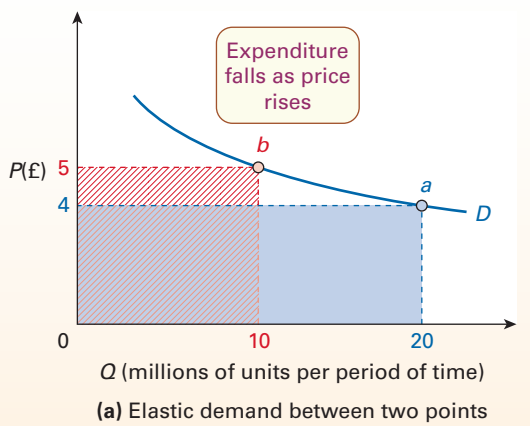
Infinitely elastic demand. This is shown by a horizontal straight line. At any price above P_1 in Figure 3.3(b) demand is zero. But at P_1 (or any price below) demand is 'infinitely' large.

This seemingly unlikely demand curve is in fact relatively common for an *individual producer*. In a perfect market, as we have seen, firms are small relative to the whole market (like the small-scale grain farmer). They have to accept the price as given by supply and demand in the *whole market*, but at that price they can sell as much as they produce. (Demand is not *literally* infinite, but as far as the firm is concerned it is.) In this case, the more the individual farmer produces, the more revenue will be earned. In Figure 3.3(b), more revenue is earned at Q_2 than at Q_1.

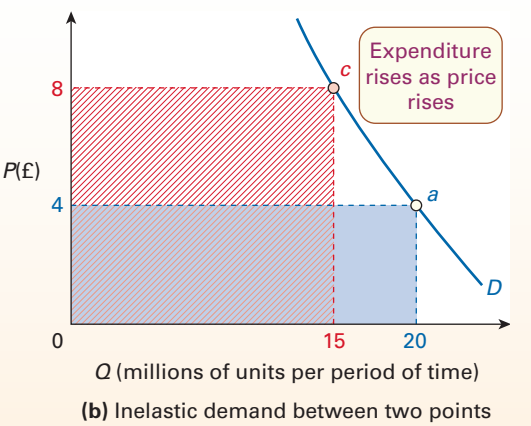
Unit elastic demand. This is where price and quantity change in exactly the same proportion. Any rise in price will be exactly offset by a fall in quantity, leaving total revenue unchanged. In Figure 3.3(c) the striped area is exactly equal to the lilac shaded area: in both cases total expenditure is £800.

You might have thought that a demand curve with unit elasticity would be a straight line at 45° to the axes.

Figure 3.2 Price elasticity of demand and total expenditure

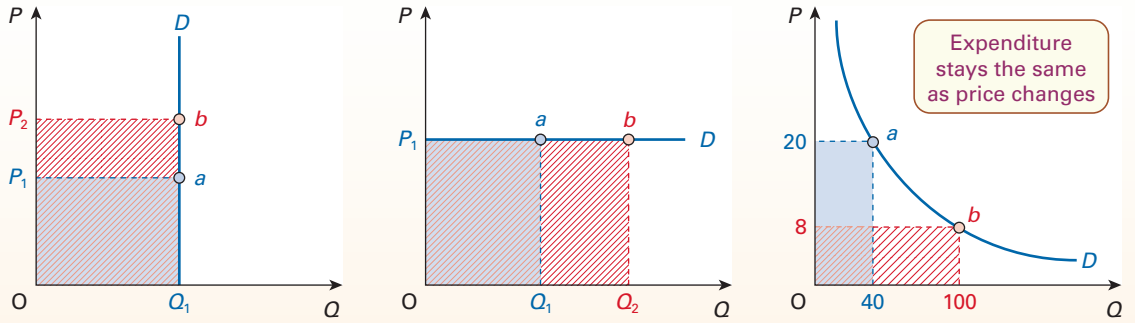


(a) Elastic demand between two points



(b) Inelastic demand between two points

Figure 3.3 Elasticity of demand: special cases



(a) Totally inelastic demand ($P\varepsilon_D = 0$)

(b) Infinitely elastic demand ($P\varepsilon_D = -\infty$)

(c) Unit elastic demand ($P\varepsilon_D = -1$)

Instead it is a curve. The reason for its particular shape is that the proportionate *rise* in quantity must equal the proportionate *fall* in price (and vice versa). As we move down the demand curve, in order for the *proportionate* change in both price and quantity to remain constant there must be a bigger and bigger *absolute* rise in quantity and a smaller and smaller absolute fall in price. For example, a rise in quantity from 200 to 400 is the same proportionate change as a rise from 100 to 200, but its absolute size is double. A fall in price from £5 to £2.50 is the same percentage as a fall from £10 to £5, but its absolute size is only half.

Pause for thought

Two customers go to the fish counter at a supermarket to buy some cod. Neither looks at the price. Customer A orders 1 kilo of cod. Customer B orders £3 worth of cod. What is the price elasticity of demand of each of the two customers?

Recap

1. The total expenditure on a product is found by multiplying the quantity sold by the price of the product.
2. When demand is price elastic, a rise in price will lead to a reduction in total expenditure on the good and hence a reduction in the total revenue of producers.
3. When demand is price inelastic, a rise in price will lead to an increase in total expenditure on the good and hence an increase in the total revenue of producers.

3.3 PRICE ELASTICITY OF SUPPLY (Pe_S)

How responsive is supply to a change in price?

When price changes, there will be not only a change in the quantity demanded, but also a change in the quantity *supplied*. Frequently we will want to know just how responsive quantity supplied is to a change in price. The measure we use is the **price elasticity of supply**.

Figure 3.4 shows two supply curves. Curve S_2 is more elastic between any two prices than curve S_1. Thus, when price rises from P_1 to P_2 there is a larger increase in quantity supplied with S_2 (namely, Q_1 to Q_3) than there is with S_1 (namely, Q_1 to Q_2). For any shift in the demand curve there will be a larger change in quantity supplied and a smaller change in price with curve S_2 than with curve S_1. Thus the effect on price and quantity of a shift in the demand curve will depend on the price elasticity of supply.

Definition

Price elasticity of supply The responsiveness of quantity supplied to a change in price.

BOX 3.2 **ADVERTISING AND ITS EFFECT ON DEMAND CURVES** CASE STUDIES & APPLICATIONS

How to increase sales and price

TC 6 p 24

When we are told that brand X will make us more beautiful, enrich our lives, wash our clothes whiter, help us attract more friends, give us a new taste sensation or allow us to save the planet, just what are the advertisers up to? 'Trying to sell the product', you may reply. In fact there is a bit more to it than this. Advertisers are trying to do two things:

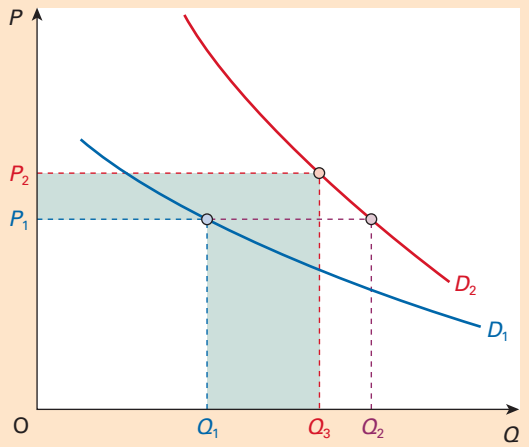
- Shift the product's demand curve to the right.
- Make it less price elastic.

This is illustrated in the diagram.

D_1 shows the original demand curve with price at P_1 and sales at Q_1. D_2 shows the curve after an advertising campaign. The rightward shift allows an increased quantity (Q_2) to be sold at the original price. If the demand is also made highly inelastic, the firm can also raise its price and still have a substantial increase in sales. Thus in the diagram, price can be raised to P_2 and sales will be Q_3, still substantially above Q_1. The total gain in revenue is shown by the shaded area.

How can advertising bring about this new demand curve?

Effect of advertising on the demand curve



Shifting the demand curve to the right

This can occur in two ways. First, if the advertising brings the product to more people's attention, then the market for the good grows and the demand curve shifts to the right. Secondly, if the advertising increases people's desire for the characteristics or attributes of a product, they will be prepared to pay a higher price for each unit purchased.

Making the demand curve less elastic

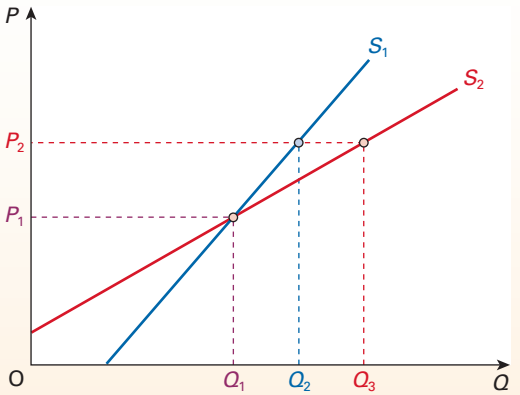
This will occur if the advertising creates a greater attachment to a product, a range of products or indeed to an organisation (brand loyalty). In other words, advertising can be used to affect the perceptions of products and services and even those of organisations, including, for example, firms, universities, charities and sports clubs.

If advertising is successful in generating a sense of uniqueness among consumers for a product it allows a firm to raise the price of the product without a significant fall in sales. If it is successful in differentiating its brand, then the firm can raise prices across its range of products, again without hitting sales. Successful advertising therefore reduces the size of the substitution effect (see page 31) since consumers have been led to believe that there are few, if any, close substitutes.

Imagine that 'Sunshine' sunflower margarine, a well-known brand, is advertised with the slogan, 'It helps you live longer'. What do you think would happen to the demand curve for a supermarket's own brand of sunflower margarine? Consider both the direction of shift and the effect on elasticity. Will the elasticity differ markedly at different prices? How will this affect the pricing policy and sales of the supermarket's own brand? What do you think might be the response of government to the slogan?

Research some recent advertising campaigns. Identify what characteristics or attributes the advertising was attempting to highlight and what perceptions it was trying to influence. Where possible, examine the impact of the campaign so far.

Figure 3.4 Two supply curves of different elasticity



The **formula for the price elasticity of supply** ($P\epsilon_S$) is: the percentage (or proportionate) change in quantity supplied divided by the percentage (or proportionate) change in price. Putting this in symbols gives:

$$P\epsilon_S = \frac{\%\,\Delta Q_S}{\%\,\Delta P}$$

In other words, the formula is identical to that for the price elasticity of demand, except that quantity in this case is quantity *supplied*. Thus if a 10 per cent rise

Definition

Formula for price elasticity of supply (arc method) ΔQ_S/average $Q_S \div \Delta P$/average P.

in price caused a 20 per cent rise in the quantity supplied, the price elasticity of supply would be:

$$20\%/10\% = 2$$

and if a 10 per cent rise in price caused only a 5 per cent rise in the quantity, the price elasticity of supply would be:

$$5\%/10\% = 0.5$$

In the first case, supply is elastic ($P\epsilon_S > 1$); in the second it is inelastic ($P\epsilon_S < 1$). Notice that, unlike the price elasticity of demand, the figure is positive (assuming that the supply curve is upward sloping). This is because price and quantity supplied change in the *same* direction.

The determinants of price elasticity of supply

The amount that costs rise as output rises. The less the additional costs of producing additional output, the more firms will be encouraged to produce for a given price rise: the more elastic supply will be.

Supply is thus likely to be elastic if firms have plenty of spare capacity, if they can readily get extra supplies of raw materials, if they can easily switch away from producing alternative products and if they can avoid having to introduce overtime working (at higher rates of pay). If all these conditions hold, costs will be little affected by a rise in output and supply will be relatively elastic. The less these conditions apply, the less elastic will supply be.

Time period

- Immediate time period. Firms are unlikely to be able to increase supply by much immediately. Supply is virtually fixed, or can vary only according to available stocks. Supply is highly inelastic.
- Short run. Over a slightly longer period of time, some inputs can be increased (e.g. raw materials) while others will remain fixed (e.g. heavy machinery). Supply can increase somewhat.
- Long run. In the long run, there will be sufficient time for all inputs to be increased and for new firms to enter the industry. Supply, therefore, is likely to be highly elastic. In some circumstances the long-run supply curve may even slope downwards (see Section 5.2).

Recap

1. Price elasticity of supply measures the responsiveness of supply to a change in price. It has a positive value.
2. Supply will be more elastic the less costs per unit rise as output rises and the longer the time period.

3.4 OTHER ELASTICITIES

How does demand respond to changes in income and to changes in the price of other goods?

Elasticities measure the responsiveness of one variable to a change in another. So far we have looked at the responsiveness of demand and supply to a change in price. But price is just one of the determinants of demand and supply. In theory, we could look at the responsiveness of demand or supply to a change in *any* one of their determinants. We could have a whole range of different types of elasticity of demand and supply.

There are two other elasticities that are particularly useful to us, and both are demand elasticities: the income elasticity of demand and the cross-price elasticity of demand.

Income elasticity of demand

The **income elasticity of demand** ($Y\epsilon_D$) measures the responsiveness of demand to a change in consumer incomes (Y).[5] It enables us to predict how much the demand curve will shift for a given change in income.

The **formula for income elasticity of demand** is the percentage (or proportionate) change in demand divided by the percentage (or proportionate) change in income. Putting this in symbols gives:

$$Y\epsilon_D = \frac{\%\Delta Q_D}{\%\Delta Y}$$

In other words, the formula is identical to that for the price elasticity of demand, except that we are dividing the change in demand by the change in income that caused it rather than by a change in price. Thus if a

Definitions

Income elasticity of demand ($Y\epsilon_D$) The responsiveness of demand to a change in consumer incomes.

Formula for income elasticity of demand The percentage (or proportionate) change in demand divided by the percentage (or proportionate) change in income: $\%\Delta Q_D \div \%\Delta Y$.

[5]Note that we use the letter Y rather than the letter I to stand for 'income'. This is normal practice in economics. The reason is that the letter I is used for 'investment'.

2 per cent rise in income caused an 8 per cent rise in a product's demand, then its income elasticity of demand would be:

$$8\%/2\% = 4$$

Pause for thought

Assume that you decide to spend a quarter of your income on clothes. What is (a) your income elasticity of demand; (b) your price elasticity of demand?

What determines income elasticity of demand? The major determinant of income elasticity of demand is the degree of 'necessity' of the good. In a developed country, the demand for luxury goods expands rapidly as people's incomes rise, whereas the demand for basic goods, such as bread, rises only a little. Thus items such as designer handbags and foreign holidays have a high income elasticity of demand, whereas items such as vegetables and socks have a low income elasticity of demand.

How do we interpret the figure for income elasticity of demand? If income elasticity of demand is positive and greater than 1, then this tells us that the share of consumers' income spent of the good increases as their income rises. If the figure is positive but less than 1, this tells us that the share of consumers' income spent on the good falls as income rises. Nevertheless, in both of these cases people demand more of the good as incomes rise.

However, the demand for some goods actually *decreases* as income rises. These are *inferior goods*, such as many of the supermarkets' value lines. As people earn more, so they switch to better quality products. Unlike **normal goods**, which have a positive income elasticity of demand, **inferior goods** have a negative income elasticity of demand.

The importance of the concept of income elasticity of demand. Income elasticity of demand is an important concept to firms considering the future size of the market for their product. If the product has a high income elasticity of demand, sales are likely to expand rapidly as national income rises, but may also fall significantly if the economy moves into recession. (See Case Study 3.3, *Income elasticity of demand and the balance of payments*, on the student website. This shows how the concept of income elasticity of demand can help us understand why so many developing countries suffered from chronic balance of trade problems in the past, but why many have seen an improvement in recent years.)

Cross-price elasticity of demand

This is often known by its less cumbersome title of **cross elasticity of demand**. It is a measure of the responsiveness of demand for one product to a change in the price of another (either a substitute or a complement). It enables us to predict how much the demand curve for the first product will shift when the price of the second product changes.

The **formula for cross-price elasticity of demand** ($C\epsilon_{Dab}$) is: the percentage (or proportionate) change in demand for good a divided by the percentage (or proportionate) change in price of good b. Putting this in symbols gives:

$$C\epsilon_{Dab} = \frac{\%\Delta Q_{Da}}{\%\Delta P_b}$$

How do we interpret the figure for cross elasticity of demand? If good b is a *substitute* for good a, a's demand will *rise* as b's price rises. In this case, cross elasticity will be a positive figure. For example, if the demand for butter rose by 2 per cent when the price of non-dairy spreads (a substitute) rose by 8 per cent, then the cross elasticity of demand for butter with respect to non-dairy spreads would be:

$$2\%/8\% = 0.25$$

If good b is *complementary* to good a, however, a's demand will *fall* as b's price rises and thus as the quantity of b demanded falls. In this case, cross elasticity of demand will be a negative figure. For example, if a 4 per cent rise in the price of bread led to a 3 per cent fall in demand for butter, the cross elasticity of demand for butter with respect to bread would be:

$$-3\%/4\% = -0.75$$

What determines cross elasticity of demand? The major determinant of cross elasticity of demand is the closeness of the substitute or complement. The closer it is, the bigger will be the effect on the first good of a change in the price of the substitute or complement, and hence the greater the cross elasticity – either positive or negative. For example, a figure of +1.169 has been estimated for the cross-price elasticity of demand for on-trade spirits

Definitions

Normal goods Goods whose demand increases as consumer incomes increase. They have a positive income elasticity of demand. Luxury goods will have a higher income elasticity of demand than more basic goods.

Inferior goods Goods whose demand *decreases* as consumer incomes increase. Such goods have a negative income elasticity of demand.

Cross-price elasticity of demand The responsiveness of demand for one good to a change in the price of another.

Formula for cross-price elasticity of demand ($C\epsilon_{Dab}$) The percentage (or proportionate) change in demand for good a divided by the percentage (or proportionate) change in price of good b: $\%\Delta Q_{Da} \div \%\Delta P_b$.

(i.e. whisky, vodka, etc.) with respect to the price of on-trade beer. This suggests they are moderately close substitutes in consumption.[6]

The usefulness of the concept of cross elasticity of demand. Firms need to know the cross elasticity of demand for their product when considering the effect on the demand for their product of a change in the price of a rival's product or of a complementary product. These are vital pieces of information for firms when making their production plans. For example, knowledge of the cross elasticity of demand for Coca-Cola® to changes in the price of Pepsi™ would allow Coca-Cola to predict the effect on its own sales if the price of Pepsi were to change.

[6]P. Meier et al. (op. cit.)

Another example of the usefulness of the concept of cross elasticity of demand is in the field of international trade and the balance of payments. A government will wish to know how a change in domestic prices will affect the demand for imports. If there is a high cross elasticity of demand for imports (because they are close substitutes for home-produced goods), and if prices at home rise due to inflation, the demand for imports will rise substantially, thus worsening the balance of trade.

TC 10 p 49

Pause for thought

Which are likely to have the highest cross elasticity of demand: two brands of coffee, or coffee and tea?

BOX 3.3 ELASTICITIES AND RELATIONSHIPS CASE STUDIES & APPLICATIONS

Where there's a relationship, there's an elasticity

Understanding elasticities and their determinants provides important insights into economic relationships. Where there are economic relationships, there are elasticities to analyse and understand. We can state the general formula for any elasticity as:

$$\epsilon_{XY} = \frac{\%\,\Delta X}{\%\,\Delta Y}$$

where the formula refers to the responsiveness of variable X to a change in variable Y (where X could be quantity supplied or demanded, and Y could be price, income, the price of substitutes, or any other determinant of demand or supply).

To illustrate the broad applicability of elasticity across economics, we consider two quite different applications. The first is from macroeconomics and relates to aggregate demand: the total demand for an economy's goods and services. The second example is from the economics of sport: the demand to watch sport.

Interest-rate elasticity of aggregate demand

Understanding the impact of interest rate changes on levels of spending is crucially important for central banks, such as the Bank of England and the European Central Bank. Central banks attempt to affect their economy's interest rates to influence spending levels. Central banks raise interest rates to dampen demand and cut interest rates to increase it.

The extent to which aggregate demand responds to changes in interest rates is known as the interest-rate elasticity of aggregate demand. The more sensitive aggregate demand is to changes in interest rates, the more interest-rate elastic is aggregate demand and the less the central bank will need to change interest rates to achieve the desired change in aggregate demand.

It is important for central banks to understand the channels through which interest rates changes influence spending. This can be done by analysing the various sectors, both at home and abroad, that generate demands for an economy's goods and services and how these demands are affected by interest rate changes. It would involve analysing, for example, the interest-rate elasticity of consumer spending and firms' investment spending, which are key components of aggregate demand (see Chapter 9).

The win-rate elasticity of demand

The demand to watch sports' events, particularly team sports like football, rugby and cricket, can be modelled in the same way as the demand for more conventional goods and services. One factor that may influence the level of demand for such events is the 'win success rate' of supporters' home teams.

In the case of team sports, some clubs may have a large proportion of so-called 'fickle' fans. These fans will only pay to watch a team that has been winning its matches; their interest quickly wanes if the team starts losing. Their attendance can be described as 'win-rate responsive' or 'win-rate elastic'. In other words, their likelihood of attending is highly dependent on the team's win rate. Other clubs may have a high proportion of 'loyal' fans. These fans are prepared to pay whether the team wins or loses. These fans are relatively unresponsive to the team's performance; their attendance is win-rate inelastic. Changes in the level of success of such clubs do little to affect attendance.

1. *Assume that the rate of inflation is forecast to rise. Of what significance might the interest-rate elasticity of aggregate demand be in determining the central bank's interest rate response?*
2. *Assume that a potential sponsor is looking to invest in one of two comparable football clubs but with different estimated win-rate elasticities of demand. How might this influence the sponsor's investment decision?*

Visit the websites of the Bank of England and the European Central Bank (ECB) and research the concept of the transmission mechanism of monetary policy. Write a short summary of what you understand by the concept and the relevance of elasticities to the mechanism. In your summary draw out similarities and differences in how each central bank describes the transmission mechanism.

Recap

1. Income elasticity of demand measures the responsiveness of demand to a change in income. For normal goods it has a positive value; for inferior goods it has a negative value.
2. Demand will be more income elastic the more luxurious the good and the less rapidly demand is satisfied as consumption increases.
3. Cross-price elasticity of demand measures the responsiveness of demand for one good to a change in the price of another. For substitute goods the value will be positive; for complements it will be negative.
4. The cross-price elasticity will be more elastic the closer the two goods are as substitutes or complements.

3.5 MARKETS AND ADJUSTMENT OVER TIME

How do markets respond over the longer term to a change in demand or supply?

The full adjustment of price, demand and supply to a situation of disequilibrium will not be instantaneous. KI 8 p 24 It is necessary, therefore, to analyse the time path which supply takes in responding to changes in demand, and which demand takes in responding to changes in supply.

Short-run and long-run adjustment

As we have already seen, the price elasticities of demand and supply vary with the time period under consideration. The reason is that producers and consumers take time to respond to a change in price. The longer the time period, the bigger the response, and thus the greater the elasticity of demand and supply.

TC 4 p 13 This is illustrated in Figures 3.5 and 3.6. In both cases, as equilibrium moves from points *a* to *b* to *c*, there is a large short-run price change (P_1 to P_2) and a small short-run quantity change (Q_1 to Q_2), but a small long-run price change (P_1 to P_3) and a large long-run quantity change (Q_1 to Q_3).

Price expectations and speculation

In a world of shifting demand and supply curves, prices do not stay the same. Sometimes they go up; sometimes they come down. TC 6 p 24 If prices are likely to change in the foreseeable future, this will affect the behaviour of buyers and sellers *now*. If, for example, it is now December and you are thinking of buying a new winter coat, you might decide to wait until the January sales, and in the meantime make do with your old coat. If, on the other hand, when January comes you see a new summer jacket in the sales, you might well buy it now and not wait until the summer for fear that the price will have gone up by then. Thus a belief that prices will go up will cause people to buy now; a belief that prices will come down will cause them to wait.

The reverse applies to sellers. If you are thinking of selling your house and prices are falling, you will want to sell it as quickly as possible. If, on the other hand, prices are rising sharply, you will wait as long as possible so as to get the highest price. Thus a

Figure 3.5 Response of supply to an increase in demand

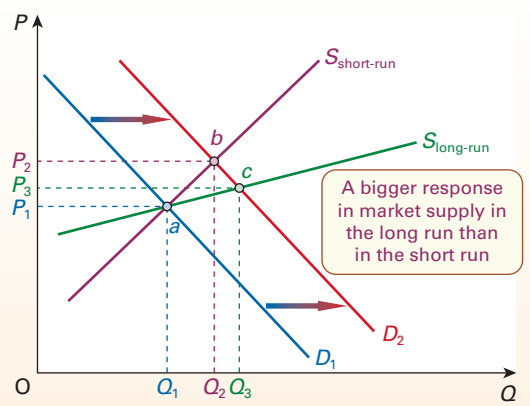
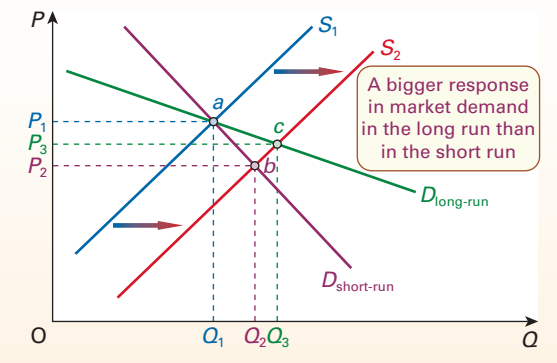


Figure 3.6 Response of demand to an increase in supply



belief that prices will come down will cause people to sell now; a belief that prices will go up will cause them to wait.

KEY IDEA 13 TC 11

People's actions are influenced by their expectations. People respond not just to what is happening now (such as a change in price), but to what they anticipate will happen in the future. Understanding the crucial role that expectations play in determining economic behaviour makes this a threshold concept – the eleventh of our fifteen.

This behaviour of looking into the future and making buying and selling decisions based on your predictions is called **speculation**. Speculation is often based on current trends in price behaviour. If prices are currently rising, people may then try to decide whether they are about to peak and go back down again, or whether they are likely to go on rising. Having made their prediction, they will then act on it. This speculation will thus affect demand and supply, which in turn will affect price. Speculation is commonplace in many markets: the stock exchange, the foreign exchange market and the housing market are three examples.

Speculation tends to be **self-fulfilling**. In other words, the actions of speculators tend to bring about the very effect on prices that speculators had anticipated. For example, if speculators believe that the price of Barclays shares is about to rise, they will buy more of them. But by doing this they will ensure that the price *will* rise. The prophecy has become self-fulfilling.

Speculation can either help to reduce price fluctuations or aggravate them: it can be stabilising or destabilising.

Stabilising speculation

Speculation will tend to have a **stabilising** effect on price fluctuations when suppliers and/or demanders believe that a change in price is only *temporary*.

Assume, for example, that there has recently been a rise in price, caused, say, by an increase in demand. In Figure 3.7(a) demand has shifted from D_1 to D_2. Equilibrium has moved from point *a* to point *b*, and price has risen from P_1 to P_2. How do people react to this rise in price? TC 4 p 13

Given that they believe this rise in price to be only temporary, suppliers bring their goods to market now, before price falls again. Supply shifts from S_1 to S_2. Demanders, however, hold back until price does fall. Demand shifts from D_2 to D_3. The equilibrium moves to point *c*, with price falling back towards P_1.

A good example of stabilising speculation occurs in agricultural commodity markets. Take the case of wheat. When it is harvested in the autumn there will be a plentiful supply. If all this wheat were to be put on the market, the price would fall to a very low level. Later in the year, when most of the wheat would have been sold, the price would then rise to a very high level. This is all easily predictable.

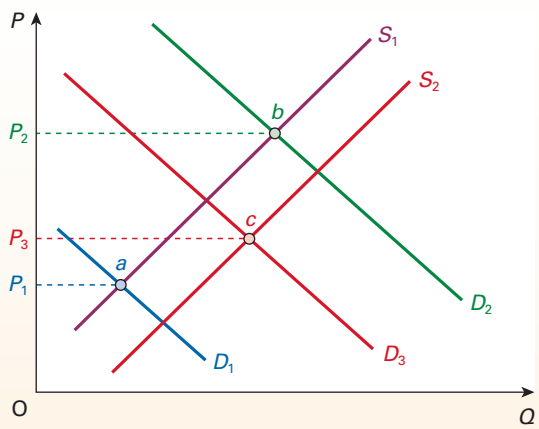
Definitions

Speculation Where people make buying or selling decisions based on their anticipations of future prices.

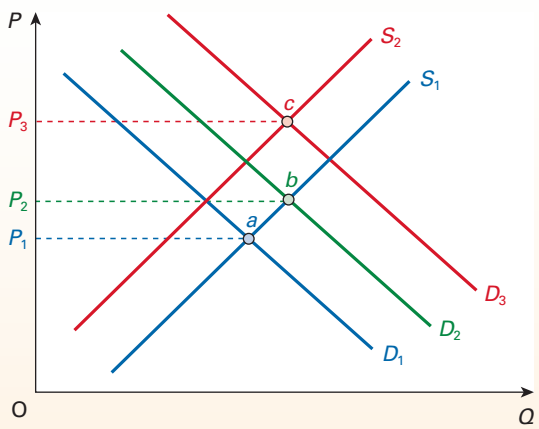
Self-fulfilling speculation The actions of speculators tend to cause the very effect that they had anticipated.

Stabilising speculation Where the actions of speculators tend to reduce price fluctuations.

Figure 3.7 Speculation (initial price rise)



(a) Stabilising speculation



(b) Destabilising speculation

So what do farmers do? The answer is that they speculate. When the wheat is harvested, they know the price will tend to fall, and so instead of bringing it all to market they put a lot of it into store. The more the price falls, the more they will put into store *anticipating that the price will later rise*. But this holding back of supplies prevents prices from falling. In other words, it stabilises prices.

Later in the year, when the price begins to rise, they will gradually release grain on to the market from the stores. The more the price rises, the more they will release on to the market *anticipating that the price will fall again by the time of the next harvest*. But this releasing of supplies will again stabilise prices by preventing them rising so much.

Rather than the farmers doing the speculation, it could be done by grain merchants. When there is a glut of wheat in the autumn, and prices are relatively low, they buy wheat on the grain market and put it into store. When there is a shortage in the spring and summer they sell wheat from their stores. In this way they stabilise prices just as the farmers did when they were the ones that operated the stores.

Destabilising speculation

Speculation will tend to have a **destabilising** effect on price fluctuations when suppliers and/or buyers believe that a change in price heralds similar changes to come.

Assume again that there has recently been a rise in price, caused by an increase in demand. In Figure 3.7(b), demand has shifted from D_1 to D_2 and price has risen from P_1 to P_2. This time, believing that the rise in price heralds further rises to come, suppliers wait until the price rises further. Supply shifts from S_1 to S_2. Demanders buy now before any further rise in price. Demand shifts from D_2 to D_3. As a result, the price continues to rise: to P_3.

Box 2.1 examined the housing market. In this market, speculation is frequently destabilising. Assume that people see house prices beginning to rise. They may well believe that this signals a boom in the housing market: that prices will go on rising. Potential buyers will thus try to buy as soon as possible before prices rise any further. This increased demand (as in Figure 3.7(b)) will thus lead to even bigger price rises. This is precisely what happened in the UK housing market in 1999–2007 and from late 2020. Conversely, in early 2008 prices started to fall; potential buyers believed that they would fall further and thus held off entering the market, leading to even bigger price falls.

A good example of a market that exhibits extreme price volatility is bitcoin. There has been massive speculation in this digital currency that was created in 2009.

Pause for thought

Draw two diagrams like Figures 3.7(a) and (b), only this time assume an initial fall in demand and hence price. The first diagram should show the effects of stabilising speculation and the second the effect of destabilising speculation.

Conclusion

In some circumstances, then, the action of speculators can help keep price fluctuations to a minimum (stabilising speculation). This is most likely when markets are relatively stable in the first place, with only moderate underlying shifts in demand and supply.

In other circumstances, however, speculation can make price fluctuations much worse. This is most likely in times of uncertainty, when there are significant changes in the determinants of demand and supply. Given this uncertainty, people may see price changes as signifying some trend. They then 'jump on the bandwagon' and do what the rest are doing, further fuelling the rise or fall in price.

Definition

Destabilising speculation Where the actions of speculators tend to make price movements larger.

Recap

1. A complete understanding of markets must take into account the time dimension.
2. Given that producers and consumers take a time to respond fully to price changes, we can identify different equilibria after the lapse of different lengths of time. Generally, short-run supply and demand tend to be less price elastic than long-run supply and demand. As a result, any shifts in *demand* or *supply* curves tend to have a relatively bigger effect on price in the short run and a relatively bigger effect on quantity in the long run.
3. People often anticipate price changes and this will affect the amount they demand or supply. This speculation will tend to stabilise prices (i.e. reduce fluctuations) if people believe that the price changes are only temporary. However, speculation will tend to destabilise prices (i.e. make price changes larger) if people believe that prices are likely to continue to move in the same direction as at present – at least for some time.

BOX 3.4 **SHORT SELLING** EXPLORING ECONOMICS

Gambling on a fall in share prices

Short selling is a form of speculation that can be very damaging to stock markets. This is where people take advantage of anticipated falls in share prices by selling shares they do not possess. How does this work?

TC 11 p 61

Assume that a share price is currently £10 per share and traders on the stock market believe that the price is about to fall. They want to take advantage of this but don't possess any. What they do is borrow shares from dealers who do own some and agree to return them on a specified date. They pay a fee for doing this. In the meantime they sell the shares on the market at the current price of £10 and wait for it to fall. They are now 'short' of the shares (i.e. they don't possess them but still owe them).

Assume that just before the agreed time comes for returning the shares the price has fallen to £8. The trader then buys the shares, returns them to the dealer who had lent them and pockets the difference of £2 (minus the fee).

Although anyone can short sell shares, it is largely traders from various financial institutions who engage in this practice. Huge bonuses can be earned from their employers if the short selling is profitable. This encourages an atmosphere of risk-taking and looking to short-term gains rather than providing long-term capital to firms.

Short selling in the banking crisis of 2008

The practice of short selling had become rife in the 2000s and added to the instability of markets, driving share prices down that were anticipated to fall. This was a particular problem in 2008, when worries about bad debts and losses in the banking sector led many traders to short sell the shares of banks and other financial institutions felt to be most at risk.

The short selling of Halifax Bank of Scotland (HBOS) shares in September 2008 was a major contributing factor to the collapse in its share price. HBOS, the UK's largest mortgage lender, had been suffering losses as a result of falling house prices and the difficulties of many house owners in keeping up with their monthly mortgage payments. The share price plummeted by over 70 per cent in the space of a few days. The fall was driven on by speculation, much of it short selling. On 17 September it was announced that HBOS would be taken over by Lloyds TSB.

Concerns about the practice of short selling driving instability in financial markets have led a number of governments – or agencies acting on their behalf – to ban the practice. In September 2008 the Financial Services Authority, the UK industry's regulator at the time, announced a four-month ban on the practice. At the same time, the US financial regulator, the Securities and Exchange Commission, announced a similar move. Both these bans were imposed for a matter of months, but Denmark held a similar policy for more than two years.

In May 2012, the EU passed a law giving the European Securities and Markets Authority (ESMA) the power to ban short selling in emergency situations: i.e. where it threatens the stability of the EU financial system. The UK government opposed the legislation, but the EU Court of Justice rejected the challenge in 2014.

Why would owners of shares, such as pension funds, lend them to short sellers rather than selling the shares themselves and then buying them back later?

Undertake desktop research on the positions taken by different countries in relation to short selling. Summarise your findings.

Definition

Short selling (or **shorting**) Where investors borrow an asset, such as shares or foreign currency; sell the asset, hoping the price will soon fall; then buy it back later and return it to the lender. Assuming the price has fallen, the short seller will make a profit of the difference (minus any fees).

3.6 MARKETS WHERE PRICES ARE CONTROLLED

What happens if the government fixes prices?

TC 7 p 26

At the equilibrium price, there will be no shortage or surplus. The equilibrium price, however, may not be the most *desirable* price. The government, therefore, may prefer to keep prices above or below the equilibrium price.

Setting a minimum (high) price

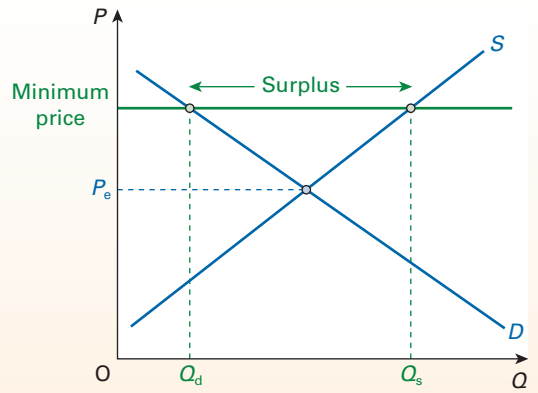
If the government sets a **minimum price** above the equilibrium (a price floor), there will be a surplus: $Q_s - Q_d$ in Figure 3.8. Price will not be allowed to fall to eliminate this surplus. The government may do this for various reasons:

- To protect producers' incomes. If the industry is subject to supply fluctuations (e.g. crops, due to fluctuations in weather) and if industry demand is price inelastic, prices are likely to fluctuate severely.

Definition

Minimum price A price floor set by the government or some other agency. The price is not allowed to fall below this level (although it is allowed to rise above it).

Figure 3.8 Minimum price: price floor



Minimum prices will prevent the fall in producers' incomes that would accompany periods of low prices.

- To create a surplus (e.g. of grains), particularly in periods of plenty, which can be stored in preparation for possible future shortages.
- To deter the consumption of particular goods. Some people may consume more of a good than is in their own self-interest because they do not fully appreciate the future costs to their health. They may also act irrationally because of self-control and addiction issues. Governments in several countries have considered introducing minimum unit pricing for alcoholic drinks (see Box 3.5).
- In the case of wages (the price of labour), minimum wages legislation can be used to prevent workers' wage rates from falling below a certain level (see Box 7.4). This may form part of a government policy on poverty and inequality.

There are various methods the government can use to deal with the surpluses associated with minimum prices.

- The government could buy the surplus and store it, destroy it or sell it abroad in other markets.
- Supply could be artificially lowered by restricting producers to particular quotas. In Figure 3.8, supply could therefore be reduced to Q_d.
- Demand could be raised by advertising, by finding alternative uses for the good, or by reducing consumption of substitute goods (e.g. by imposing taxes or quotas on substitutes, such as imports).

One of the problems with minimum prices is that firms with surplus on their hands may try to evade the price control and cut their prices.

Another problem is that high prices may cushion inefficiency. Firms may feel less need to find more efficient methods of production and cut their costs if their profits are being protected by the high price. Also the high price may discourage firms from producing alternative goods which they could produce more efficiently or which are in higher demand, but which nevertheless have a lower (free-market) price.

TC 6 p 24

Pause for thought

Draw a supply and demand diagram with the price of labour (the wage rate) on the vertical axis and the quantity of labour (the number of workers) on the horizontal axis. What will happen to employment if the government raises wages from the equilibrium to some minimum wage above the equilibrium?

Setting a maximum (low) price

If the government sets a **maximum price** below the equilibrium (a price ceiling), there will be a shortage: $Q_d - Q_s$ in Figure 3.9. Price will not be allowed to rise to eliminate this shortage. The government may set maximum prices to prevent them rising above a certain level. This will normally be done for reasons of fairness. For example, in the UK, the energy regulator, Ofgem, sets a price cap for electricity and gas.

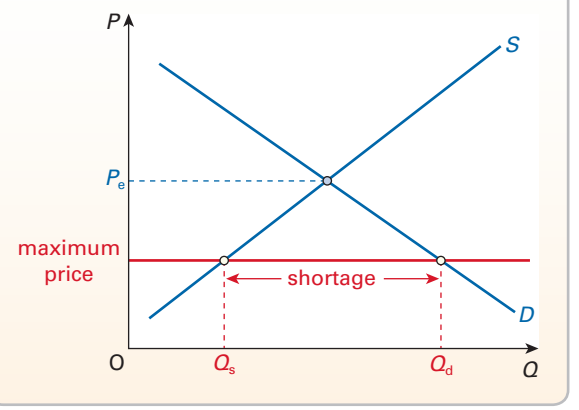
The resulting shortages, however, create further problems. If the government merely sets prices and does not intervene further, the shortages are likely to lead to the following:

- *'First-come, first-served' basis.* This is likely to lead to queues developing outside shops, or websites crashing if people try to purchase the good online. To try to deal with these issues firms may have to adopt waiting lists. Queues have been a common feature of life in Venezuela and Argentina where the governments have

Definition

Maximum price A price ceiling set by the government or some other agency. The price is not allowed to rise above this level (although it is allowed to fall below it).

Figure 3.9 Maximum price: price ceiling



BOX 3.5 **A MINIMUM UNIT PRICE FOR ALCOHOL** CASE STUDIES & APPLICATIONS

A way of reducing alcohol consumption

A market where the use of minimum pricing has been extensively discussed is that of alcohol. In early 2010, the UK House of Commons Health Select Committee[1] proposed the introduction of a form of minimum pricing. Known as a 'minimum unit price' (MUP), the price floor is set according to the alcoholic content of the drink rather than the simple volume of liquid. The Select Committee report suggested that a MUP of 50p would save more than 3000 lives per year.

In March 2012, the UK government announced its intention to introduce a MUP following a period of consultation with industry stakeholders. The Scottish government passed legislation for the introduction of a 50p MUP in June 2012.

What impact would a 50p MUP have on the price of alcoholic drinks in the UK? The table provides some examples across a range of products available in supermarkets.

Effect of a 50p MUP on the price of various alcoholic drinks

Product	Volume	Strength (% abv)	Units of alcohol	Current price	Minimum price
Cheap strong spirits	70 cl	50.0	35.00	£12.00	£17.50
Cheap spirits	70 cl	37.5	26.25	£10.00	£13.13
Cheap wine	750 ml	13.0	9.75	£3.99	£4.88
Cheap strong cider	3 litres	7.5	22.50	£3.50	£11.25
Lager (4 pack)	440 ml × 4	5.0	2.20	£4.50	£4.40
Alcopop	70 cl	4.0	2.80	£3.00	£1.40

Minimum unit prices versus higher taxes

What are the advantages of introducing a MUP rather than simply increasing tax and duties on alcohol?

Evidence from researchers at Sheffield University[2] indicates that it is a more effective way of targeting heavy drinkers, especially those who buy strong cheap drinks, such as white cider, in supermarkets. Economic simulations undertaken to forecast the impact of a 50p MUP found that the policy would have a very small impact on people who consume moderate amounts of alcohol: i.e. men who drink fewer than 22 units per week and women who drink fewer than 15 units. On average, this group responds by drinking only 3 fewer units per year, equivalent to one pint of strong beer.

The simulations suggest however that the impact of the policy on heavy drinkers would be much stronger. Men who drink over 50 units per week and women who drink over 35 units respond on average by consuming 134 fewer units per year. The impact on heavy drinkers in the poorest 20 per cent of households is even greater still. They respond by consuming 372 fewer units per year.

One disadvantage of a MUP is that any extra revenue goes to the retailers rather than to the government. Research by the Institute of Fiscal Studies,[3] suggests it would reduce competition and lead to windfall profits for the alcohol and retail industry. They argue that reforming taxes on alcohol is a more effective way of targeting heavy drinkers.

In July 2013, the UK government announced that it was not going ahead with its proposed MUP policy arguing that the evidence did not conclusively demonstrate that it would reduce problem drinking without penalising all those who drink responsibly.[4] The government instead introduced a minimum price at a much lower level. In May 2014, it became illegal for firms to sell alcoholic drinks at prices below the amount of duty and VAT levied on them.

The introduction of a MUP in Scotland was delayed by a series of legal cases brought by some parts of the alcohol industry on the grounds that it constituted a constraint on trade. In November 2017 the UK supreme court confirmed that the MUP legislation was lawful. A short period of public consultation then followed before the introduction from May 2018 of a MUP of 50 pence.

1. *Draw a diagram to illustrate the likely impact of setting a minimum price based on duty and VAT levels.*
2. *Explain how price elasticity of demand determines the impact of a 50p MUP on the consumption of alcohol.*

Undertake desktop research to compare taxation on alcohol across different countries. In summarising your findings distinguish between different types of taxation and different rates of taxation.

[1] *Alcohol: First Report of Session 2009–10*, House of Commons Health Committee (January 2010).

[2] *Minimum pricing for alcohol effectively targets high risk drinkers, with negligible effects on moderate drinkers with low incomes*, The University of Sheffield News (February 2014).

[3] *Reforms to alcohol taxes would be more effective than minimum unit pricing*, Institute for Fiscal Studies (March 2013).

[4] *Next steps following the consultation on delivering the Government's alcohol strategy*, Home Office (July 2013) p3.

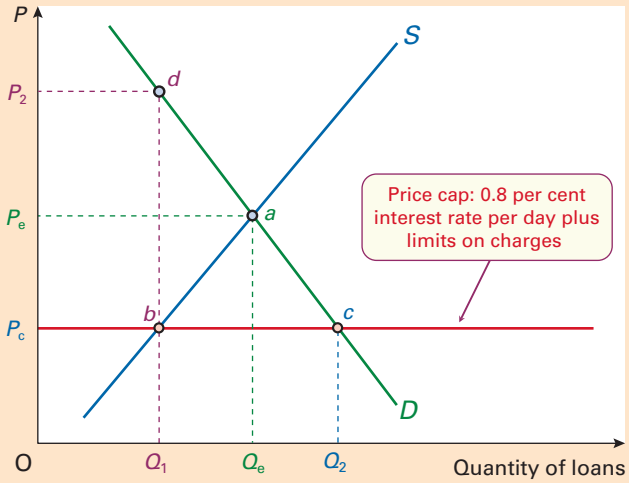
kept the prices of many basic goods below the level necessary to equate demand and supply.

- *Random ballot*. The seller puts the name of every customer willing to pay the maximum price into a random draw. Only those who are lucky enough to have their name drawn receive the good.
- *Favoured customers*. This could be the seller's friends, family and/or regular customers.
- *A measure of merit*. For example, the number of students who want a place on a particular course at a university and are willing to pay the tuition fee may exceed the number of places available.

BOX 3.6 UK PAYDAY LOAN CAP

Capping the cost of short-term credit

The impact of a maximum price on the payday loan market



A recent example of maximum pricing in the UK is in the market for high-cost, short-term credit (HCSTC). Otherwise known as the 'payday loan market', this is where customers typically borrow a few hundred pounds for a short time, usually two to four weeks. One common reason given for this type of borrowing is the need to pay an unexpected bill. The expectation is that the loan will be repaid when the borrower receives the next pay cheque.

What is the price of borrowing (i.e. buying money) in the credit market? It is the interest rate plus any other fees charged by the lender. If the market is competitive and unregulated then this price is determined by the interaction of the demand and supply. This is illustrated at point *c* in the diagram at the equilibrium price P_e.

The payday loan market grew dramatically after the financial crisis of 2007–8 and reached its peak in 2013 when the Financial Conduct Authority (FCA) estimated that approximately 400 firms made 10 million loans with a market value of £2.5 billion. However, as the market grew, concerns about the level of interest rates were expressed. For example, a borrower would typically have to pay around £38 for a 30-day loan of £100. This was an interest rate of 1.26 per cent per day or 4670 per cent per annum! Similar concerns were also expressed about the size of penalty charges imposed on borrowers when they failed to meet repayment deadlines.

In an attempt to deal with these issues, the FCA announced a series of regulations in July 2014.[1] They came into effect in January 2015 and were as follows:[2]

[1] *FCA proposes price cap for payday lenders*, Financial Conduct Authority press release (15/07/2014).

[2] *Cracking down on payday lenders will hurt the poor*, Institute for Economic Affairs (July 2013).

In this instance, the university may allocate places to those who achieve the highest grades.

- *A rule or regulation*. State schools facing excess demand often allocate places based on the distance children live from the school (in this case the maximum price is zero). Preference is also given to applicants who have an older sibling already at the school.

None of the above may be considered to be fair since some people may be forced to go without. Therefore, the government may adopt a system of **rationing** – a system commonly used in wartime as a way of coping with shortages. People could be issued with a set number of coupons for each item rationed.

A major problem with maximum prices is likely to be the emergence of **illegal markets**, which are sometimes referred to as **underground** or **shadow markets** (see Case Study 3.7 on the student website), where customers, unable to buy enough in legal markets, may well be prepared to pay very high prices: prices above P_e in Figure 3.9.

Definitions

Rationing Where the government restricts the amount of a good that people are allowed to buy.

Illegal or underground or shadow markets Where people ignore the government's price and/or quantity controls and sell illegally at whatever price equates illegal demand and supply.

CASE STUDIES & APPLICATIONS

- The interest rate was capped at 0.8 per cent per day: i.e. a maximum price was introduced. This means that the maximum amount a lender can charge for a 30-day loan of £100 is £24 (assuming it is repaid on time).
- Charges for late repayments of loans were limited to a maximum amount of £15. The interest rate on outstanding debts was also capped at 0.8 per cent per day.
- The total amount borrowers could be charged, even if they defaulted, was limited to double the amount they borrowed.

Impact of price cap

What is the impact of the maximum price on a competitive market if it is set below its equilibrium level? This is also illustrated in the diagram. With a price cap of P_c, the simple demand and supply model predicts that the quantity of the loans supplied falls from Q_e to Q_1 while the quantity demanded increases from Q_e to Q_2. There is a shortage of loans of $Q_2 - Q_1$ or $c - b$.

The FCA anticipated that the number of loans would fall, causing many lenders to go out of business. It forecast that, out of the 400 businesses registered as official payday lenders prior to the regulation, only four (one high street and three online) would survive. In reality, the reduction has been far smaller. In December 2016, the FCA announced that the number had fallen from 400 to 144.

The demand and supply model also predicts that the imposition of a maximum price gives an opportunity for the development of an illegal market. If supply is restricted to Q_1, the market clearing price would now be P_2 (point d), where demand also equals Q_1. If the legal price, therefore, is only P_c, there will be many people who would be willing to pay far more.

Let us say that an individual borrower is prepared to pay P_2. Individual payday lenders may be willing to break the law and charge a price above P_c, but below P_2, depending on their position on the supply curve. Thus both the borrower and the lender would be better off if they made a loan agreement at any price between the price the borrower is willing to pay (assuming borrowers are rational) and the price the lender is willing to accept.

However, the maximum price legislation prevents this seemingly mutually beneficial transaction from taking place legally. If potential borrowers are willing to go to the unofficial market they may find illegal lenders (loan sharks) willing to lend them money at an interest rate above the capped level of 0.8 per cent per day but below the maximum rate they are willing to pay.

In November 2016, the FCA announced that it was launching a review of its payday market regulations and would look for any evidence of consumers turning to illegal money lenders because of the price cap.[3] The review published in July 2017 concluded that its regulation had delivered substantial savings to consumers worth £150 million per year with a consequent reduction in the numbers incurring debt problems. Consequently, it decided to leave the loan price cap in place. A subsequent review in 2020 found that, despite the benefits of the price cap, firms were still lending to customers beyond levels which they could sensibly manage and causing customer harm as a result.[4]

1. *Using the concept of elasticity, explain why the impact of a maximum price, such as the one introduced by the FCA, may be different in the economic long run as opposed to the short run.*
2. *Discuss the economic rationale for introducing a maximum price into the market for payday lending.*

Using the Insolvency Statistics produced by the Insolvency Service, write a short briefing note on the recent patterns in insolvencies of individuals in the UK.

[3] *FCA launches call for input on high-cost credit and overdrafts*, FCA press release (29/11/2016).

[4] 'Relending by high-cost lenders', *Multi-firm reviews*, FCA (6/8/2020).

Another problem is that maximum prices reduce the quantity produced of an already scarce commodity. For example, artificially low prices in a famine are likely to reduce food supplies: if not immediately, then at the next harvest, because of less being sown. In many developing countries, governments control the price of basic foodstuffs in order to help the urban poor. The effect, however, is to reduce incomes for farmers, who are then encouraged to leave the land and flock into the ever-growing towns and cities.

TC 7 p 26

To minimise these types of problem the government may attempt to reduce the shortage by encouraging supply: by drawing on stores, by direct government production, or by giving subsidies or tax relief to firms. Alternatively, it may attempt to reduce demand: by the production of more alternative goods (e.g. home-grown vegetables in times of war) or by controlling people's incomes.

Another example of maximum prices is where the government imposes rent controls in an attempt to make rented accommodation more affordable. Here the 'price' is the rent people are charged. The danger of this policy is that it will create a shortage of rental property. The policy is examined in Case Study 3.8 on the student website.

Pause for thought

Many governments intervene in the market for private rented accommodation including setting a maximum rent payable by tenants. Consider the arguments for and against rent controls.

BOX 3.7 **THE EFFECT OF IMPOSING TAXES ON GOODS** **EXPLORING ECONOMICS**

Who ends up paying?

Another example of government intervention in markets is the imposition of taxes on goods. These indirect taxes, as they are called, include taxes such as value added tax (VAT) and excise duties on cigarettes, petrol and alcoholic drinks.

These taxes can be a fixed amount per unit sold – a 'specific tax'. An example is the tax on petrol, which is set at so much per litre. Alternatively, they can be a percentage of the price, or value added, at each stage of production – an '*ad valorem* tax'. An example is VAT.

When a tax is levied on a good or service, this has the effect of shifting the supply curve upwards by the amount of the tax (see diagram (a)). In the case of a specific tax, it will be a parallel shift, since the amount of the tax is the same at all prices.

But why does the supply curve shift upwards by the amount of the tax? In diagram (a), the supply curve (S) shows that to supply Q_1, producers need to receive a price of P_1. Now a tax is levied of an amount shown by the arrow. For producers to continue receiving P_1 and hence producing Q_1, the price charged to consumers has to be P_1 + tax. Thus the new supply curve is shown by the red line S_T.

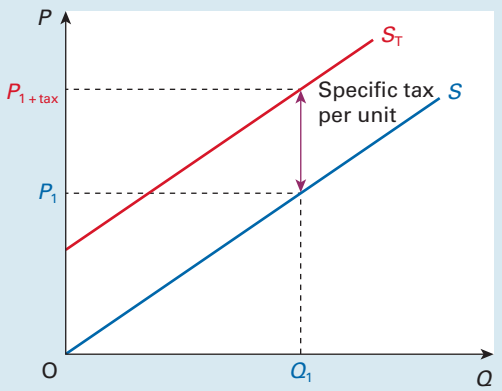
The incidence of the tax

What will be the effect of the tax on the price and the quantity sold? This is illustrated in diagram (b). Before the tax is imposed, Q_1 units are sold at a price of P_1. The effect of the tax is to shift the supply curve to S_T. Price rises to P_2 and quantity falls to Q_2.

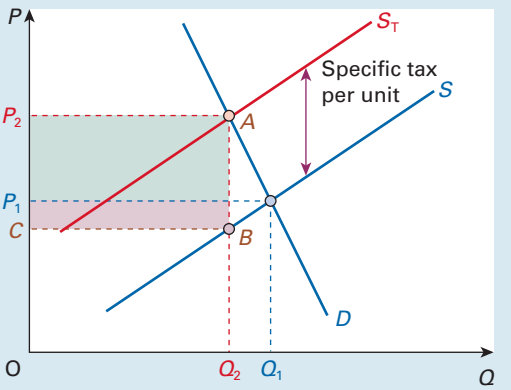
Notice, however, that price does not rise by the full amount of the tax, because the demand curve is downward sloping. The amount of the tax is $P_2 - C$, whereas the price increase is only $P_2 - P_1$. Thus the burden or incidence of such taxes is distributed between consumers and producers.

Consumers pay to the extent that price rises. Producers pay to the extent that this rise in price is not sufficient to cover the tax.

(a) Effect of a tax on the supply curve



(b) Incidence of tax



We can also show in diagram (b) the revenue the government receives from the tax. The tax per unit is $P_2 - C$ and the quantity sold is Q_2 (shown by the distance CB). Thus the revenue raised is the total area of the shaded rectangle, P_2ABC.

The rise in price from P_1 to P_2 multiplied by the number of goods sold (Q_2) (the green area) is the amount of the tax passed on to consumers and thus represents the consumers' share of the tax. The remainder (the lilac area) is the producers' share. This is the amount by which the producers' net price is below the original price (i.e. $P_1 - C$) multiplied by Q_2.

Elasticity and the incidence of taxation

It is easy to see that the less elastic demand is:

- the less quantity sold falls and hence the greater the revenue the government raises from the tax;
- the more price rises and hence the greater the proportion of the tax paid by consumers.

Thus in most countries, cigarettes, petrol and alcohol are the major targets for indirect taxes. Demand for each of them is high and relatively inelastic, and so taxes on them raise a lot of revenue and do not curb demand significantly. Indeed, in the UK, fuel duty (a specific tax) and VAT (an *ad valorem* tax), together account for between 60 and 75 per cent of the cost of petrol, depending on the price of petrol.

What this does mean, however, is that any cut in tax will be largely or completely passed on to the consumer. This was the case when the UK government reduced fuel duty by 5p per litre in March 2022 to help reduce the rise in fuel prices resulting from soaring oil prices in the wake of the Ukraine war.

Demand tends to be more elastic in the long run than in the short run. Assume that a tax is imposed on a good that was previously untaxed. How will the incidence of this tax change as time passes?

Recap

1. The government may fix minimum or maximum prices. If a minimum price is set above the equilibrium price, a surplus will result. If a maximum price is set below the equilibrium price, a shortage will result.
2. Minimum prices are set as a means of protecting the incomes of suppliers or creating a surplus for storage in case of future reductions in supply. If the government is not deliberately trying to create a surplus, it must decide what to do with it.
3. Maximum prices are set as a means of keeping prices down for the consumer. The resulting shortage will cause queues, waiting lists or the restriction of sales by firms to favoured customers. Alternatively, the government could introduce a system of rationing. If it does, then underground markets are likely to arise. This is where goods are sold illegally above the maximum price.

QUESTIONS

1. Draw a diagram with two supply curves, one steeply sloping and one gently sloping. Ensure that the two curves cross. Draw a demand curve through the point where they cross and mark the equilibrium price and quantity. Now assume that the demand curve shifts to the right. Show how the shape of the supply curve will determine just what happens to price and quantity.
2. Which of the following will have positive signs and which will have negative ones? (a) price elasticity of demand; (b) income elasticity of demand (normal good); (c) income elasticity of demand (inferior good); (d) cross elasticity of demand (with respect to changes in price of a substitute good); (e) cross elasticity of demand (with respect to changes in price of a complementary good); (f) price elasticity of supply.
3. Demand for oil might be relatively elastic over the longer term, and yet it could still be observed that over time people consume more oil (or only very slightly less) despite rising oil prices. How can this apparent contradiction be explained?
4. Is the demand curve for a particular football club's season tickets the same as that facing any other football club? Explain.
5. Assume that a football club has the following demand curve for season tickets: $Q_d = 50\,000 - 50P$.
   a. What quantity of season tickets is demanded when season tickets are free ($P = 0$)?
   b. At what price do season tickets cease to be bought ($Q = 0$)?
   c. Using the information from (a) and (b) sketch the demand curve for the club.
   d. What amount of season tickets is demanded at £600 and £700? What revenue is generated from season ticket sales at each price?
   e. Using the average (mid-point) formula in Box 3.1 estimate the price elasticity between £600 and £700.
   f. What can we say about the price elasticity of demand between the points on the demand curve corresponding to season ticket prices of £600 and £700?
   g. Between what price range does raising season ticket prices increase revenues from season ticket sales?
6. How might a firm set about making the demand for its brand less elastic?
7. Assuming that a firm faces an inelastic demand and wants to increase its total revenue, in what direction should it change its price? Is there any limit to how far it should go in changing its price in this direction?
8. Why are both the price elasticity of demand and the price elasticity of supply likely to be greater in the long run?
9. Which are likely to have the highest cross elasticity of demand: two brands of coffee, or coffee and tea? Explain.
10. Redraw both diagrams in Figure 3.7, only this time assume that it was an initial shift in supply that caused price to change in the first place.
11. What are the advantages and disadvantages of speculation from the point of view of (a) the consumer; (b) firms?
12. Assume that the (weekly) market demand and supply of tomatoes are given by the figures shown below:

Price (£ per kilo)	4.00	3.50	3.00	2.50	2.00	1.50	1.00
Q_d (000 kilos)	30	35	40	45	50	55	60
Q_s (000 kilos)	80	68	62	55	50	45	38

   a. What are the equilibrium price and quantity?
   b. What will be the effect of the government fixing a minimum price of (i) £3.00 per kilo; (ii) £1.50 per kilo?
   c. Suppose that the government paid tomato producers a subsidy of £1.00 per kilo. (i) Give the new supply schedule. (ii) What will be the new equilibrium price? (iii) How much will this cost the government?
   d. Alternatively, suppose that the government guaranteed tomato producers a price of £2.50 per kilo. (i) How many tomatoes would it have to buy in order to ensure that all the tomatoes produced were sold? (ii) How much would this cost the government?
   e. Alternatively, suppose it bought all the tomatoes produced at £2.50. (i) At what single price would it have to sell them in order to dispose of the lot? (ii) What would be the net cost of this course of action?
13. Think of two things that are provided free. In each case, identify when and in what form a shortage might occur. In what ways are/could these shortages be dealt with? Are they the best solution to the shortages?
14. Think of some examples where the price of a good or service is kept below the equilibrium. In each case consider the advantages and disadvantages of the policy.

Pearson | MyLab | Economics

Log on to MyLab Economics and complete the studyplan exercises for this chapter to see how much you have learnt and where you need to revise most. Make sure you access all the supporting textbook resources, including the online workbook, newsblog, audio animations, guided solutions and ebook.

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

3.1 **Shall we put up our price?** This uses the concept of price elasticity of demand to explain why prices are higher where firms face little or no competition.

3.2 **Any more fares?** Pricing on the buses: an illustration of the relationship between price and total revenue.

3.3 **Income elasticity of demand and the balance of payments.** This examines how a low income elasticity of demand for the exports of many developing countries can help to explain their chronic balance of payments problems.

3.4 **Elasticities of demand for various foodstuffs.** An examination of the evidence about price and income elasticities of demand for food in the UK.

3.5 **What we pay to watch sport.** Consideration of the demand function for season tickets to watch spectator sports like football.

3.6 **The role of the speculator.** This assesses whether the activities of speculators are beneficial or harmful to the rest of society.

3.7 **Underground or shadow markets.** How underground markets can develop when prices are fixed below the equilibrium.

3.8 **Rent control.** An analysis of the potential impact of rent controls on the private rental market.

3.9 **Agricultural subsidies.** This considers who gains and who loses from the use of subsidies on the production of agricultural products.

3.10 **The CAP and the environment.** This case shows how the system of high intervention prices had damaging environmental effects. It also examines the more recent measures the EU has adopted to reverse the effects.

3.11 **The Common Agricultural Policy of the EU.** This case study looks at the various forms of intervention in agriculture that have been used in the EU. It looks at successes and problems and at various reforms that have been introduced.

3.12 **The fallacy of composition.** An illustration of how something that applies to an individual may not apply to a larger group. The example is taken from agriculture.

3.13 **The cobweb model.** An outline of the theory that explains price fluctuations in terms of time lags in supply.

3.14 **Dealing in futures markets.** How buying and selling in futures markets can reduce uncertainty.

WEB APPENDIX

3.1 **Point elasticity.** An alternative way of measuring elasticity.

Chapter 4

The demand decision

In this chapter we take a more detailed look at consumer demand. If we had unlimited income and time we would not have to be careful with our money. In the real world, however, we have to make choices about what to buy. You may have to choose between buying textbooks and going to a festival, between a new pair of jeans and a meal out, between saving for a car and having more money to spend on everyday items.

We begin by exploring the concept of the 'rational consumer'. What does rationality mean in the context of consumption behaviour? How do rational consumers allocate their scarce income across all the goods they might like to consume?

In Section 4.2 we look at the problem of making rational choices when benefits occur over a period of time, as is the case with durable goods, or later, as is the case with goods where there is a waiting list or where orders take some time to deliver.

In Section 4.3 we look at consumption decisions when consumers are faced with uncertainty. When we buy goods or services there is often the risk that the payoffs (the costs and benefits) will not turn out as we had expected. We conclude this section by looking at why consumers may take out insurance to safeguard against uncertainty and at the behaviour of the firms providing insurance.

In the final part of the chapter, we look at insights from behavioural economics into how people actually behave when buying goods and services. This field of economics draws on disciplines such as psychology, neuroscience and sociology. Behavioural ideas can be applied to examine, for example, why people often do not weigh up the costs and benefits of all possible purchases, but instead behave impulsively or according to simple rules of thumb.

After studying this chapter, you should be able to answer the following questions:

- What do economists mean by the term 'utility'?
- At what point does the rational consumer cease to consume additional units of a good or service?
- How do rational consumers allocate their scarce budget across a range of consumption possibilities?
- How can the traditional approach to modelling consumer choices be extended to reflect circumstances when costs and benefits extend over several time periods?
- Why are people prepared to pay for insurance?
- How do people in practice make decisions about buying and selling, when lacking information and when they can be influenced by emotion, impulse and habit?
- What is behavioural economics and how does it differ from traditional economics when applied to consumer decisions?
- How can the framing or presentation of choices or their outcomes affect the decisions that consumers make?

4.1 CONSUMER CHOICE

Analysing the rational consumer

In Chapter 1 we saw that rational choices involve weighing up *marginal costs* and *marginal benefits*: weighing up the costs and benefits of doing a little bit more or a little bit less of a specific activity. Economists refer to the **rational consumer** as a person who attempts to get the best value for money from his or her purchases, given a limited income. Thus the rational consumer tries to ensure that the benefits of a purchase are worth the expense.

TC 2 p 9

TC 1 p 7

Sometimes we may act 'irrationally'. We may buy goods impetuously or out of habit. This is something we examine in Section 4.4. For now, let us assume that people behave rationally.

Two words of warning before we go on. First, don't confuse irrationality and ignorance. Although consumers may behave rationally, that does not mean that they have perfect information. Have you ever been disappointed after buying something? Perhaps it was not as good as you had expected from an advert? Or perhaps you found later that you could have bought an alternative more cheaply? Perhaps a holiday did not turn out to be as good as the website led you to believe. This is a problem of ignorance rather than irrationality.

Second, the term 'rational' does not imply any approval of the decision involved. It is simply referring to behaviour that is consistent with your own particular goals: behaviour directed to getting the most out of your limited income. People may disapprove of the things that others buy – their clothes, junk food, lottery tickets – but as economists we are not making judgements about people's goals. We can, however, look at the implications of people behaving rationally in pursuit of those goals. This is what we are doing when we examine rational consumer behaviour: we are looking at its implications for consumer demand.

Definitions

Rational consumer A person who weighs up the costs and benefits to him or her of each additional unit of a good purchased.

Principle of diminishing marginal utility As more units of a good are consumed, additional units will provide less additional satisfaction than previous units.

Marginal utility (MU) The extra satisfaction gained from consuming one extra unit of a good within a given time period.

Pause for thought

If you buy something in the shop on the corner when you know that the same item could have been bought more cheaply two miles up the road in the supermarket, is your behaviour irrational? Explain.

Utility and the rational consumer

Consumption and utility

When you buy something, it's normally because you want it. You want it because you expect to get pleasure, satisfaction or some other sort of benefit from it. This applies to everything from cups of coffee, to train journeys, to music downloads, to jeans, to insurance. Economists use the term 'utility' to refer to the benefit we get from consumption. Box 4.1 considers how utility may be affected by the characteristics of the products being consumed.

Clearly, the nature and amount of utility that people get vary from one product to another, and from one person to another. But there is a simple rule that applies to virtually all people and all products. As you consume more of a product, and thus become more satisfied, so your desire for additional units of it will decline. Economists call this rule the **principle of diminishing marginal utility**.

For example, the second cup of tea in the morning gives you less additional satisfaction than the first cup. The third cup gives less still. We call the additional utility you get from consuming an extra unit of a product the **marginal utility** (**MU**). Therefore, the rule says that the marginal utility will fall as we consume more of a product over a given period of time.

KEY IDEA 14

The principle of diminishing marginal utility. **The more of a product a person consumes over a given period of time, the less will be the additional utility gained from one more unit.**

At some level of consumption, your total utility will be at a maximum. No extra satisfaction can be gained by the consumption of further units within that period of time. Thus marginal utility will be zero. Your desire for tea may be fully satisfied at seven cups per day. An eighth cup will yield no extra utility and may even give you displeasure (i.e. negative marginal utility).

There is a problem, however, with the concept of marginal utility. How can it be measured? After all,

Pause for thought

Are there any goods or services where consumers do not experience diminishing marginal utility?

we cannot get inside each other's heads to find out just how much pleasure we are getting from consuming a product!

One way round the problem is to measure marginal utility in *money terms*: in other words, the amount that a person would be prepared to pay for one more unit of a product. Thus if you were prepared to pay 70p for an extra packet of crisps per week, then we would say that your marginal utility from consuming it is 70p. As long as you are prepared to pay more than or the same as the actual price, you will buy an extra packet. If you are not prepared to pay that price, you will not.

Utility and an individual's demand curve

We can now see how utility relates to a downward-sloping demand curve. We will focus first on an individual consumer's demand and then on the whole market demand.

Rational consumers try to get as much value from their money as possible. By measuring utility in money terms, marginal utility becomes the amount of money a person would be prepared to pay to obtain one more unit. This means that if an additional unit of a good is worth more to you than the price (P) you have to pay ($MU > P$), you will buy more of it. If it is worth less to you than it costs ($MU < P$), you will not buy it or, if you are already buying some, you will buy less.

TC 2 p 9

Let us take the case where initially the price of a good exactly equals marginal utility ($MU = P$). Assume that the price now falls. As it does so, it will be worth buying extra units. You will buy more because the price is now below the amount you are prepared to pay: i.e. price is less than your marginal

BOX 4.1 **SATISFACTION AND CONSUMER DEMAND** **EXPLORING ECONOMICS**

Identifying the benefit drivers

In consuming goods and services, consumers aim to derive some form of benefit or satisfaction. Economists refer to satisfaction as utility. This utility is important because it affects consumers' preparedness to pay for goods and services. Therefore, it is important that we understand how consumers derive their utility. In other words, what are the *sources* of satisfaction from consumption? What factors influence or affect the satisfaction from these sources? Can new sources of satisfaction be created or existing ones changed?

Product attributes

Kelvin Lancaster[1] popularised the idea that individuals derive their utility from the *characteristics* or *attributes* of goods and services.

To illustrate this, consider Sarah's preparedness to pay for cappuccinos. It might be that her utility is derived from outputs such as 'the quenching of thirst', 'revival' and 'flavour'. Perhaps, if it is being consumed in a particular coffee shop, her utility is influenced not only by the characteristics of the coffee itself but by the characteristics of the environment in which the drink is being consumed.

The characteristics of products, therefore, affect consumers' preparedness to pay. These characteristics are sometimes referred to as **benefit drivers**. They might relate, for example, to the physical characteristics of a product, such as its performance, quality and ease of use, or to other factors, such as the speed of product delivery, after-sales service and reputation of the seller.

In complex products, such as housing, there can be several benefit drivers. These may appeal differently to different consumers. For instance, some groups of buyers might be prepared to pay more for a property whose location has good transport connections, while others might be attracted by the proximity to particular schools.

An analysis of benefit drivers can help in understanding better consumer choice. When we go shopping and compare one product with another, it can be the differences in the features of the various products, along with price, that determine which we end up buying.

An analysis of benefit drivers can be important to firms in understanding what consumers might value about them and their products.

The characteristics approach to consumer choice is analysed in Case Study 4.1 on the student website.

1. *Make a list of possible benefit drivers when purchasing: (i) housing; (ii) season tickets to watch live sport; (iii) kilo bags of sugar.*
2. *Why are firms interested in benefit drivers?*

Undertake research into the characteristics of housing or cars that are used when promoting and advertising these products. Summarise your findings and discuss their relevance as benefit drivers and whether other characteristics might also be relevant benefit drivers.

Definition

Benefit drivers Attributes of a good or service that affect the benefit or utility from consuming it.

[1]K. Lancaster, 'A new approach to consumer theory', *The Journal of Political Economy*, 74 (1996), 132–57.

utility. In other words, there will be an excess of utility from the consumption of another unit over the price paid. This excess is known as **marginal consumer surplus** (*MCS*). The extent of your *MCS* is simply the difference between your marginal utility from consuming another unit, i.e. your preparedness to pay, and the price you actually pay: $MCS = MU - P$.

KI 14 p 72

But as you buy more, your marginal utility from consuming each extra unit will get less and less. This is the principle of diminishing marginal utility.

How many extra units then do you buy? You will stop when the marginal utility has fallen to the new lower price of the good: when $MU = P$. At this point there is no longer any additional consumer surplus to be gained and **total consumer surplus** (*TCS*) has therefore been maximised. The total consumer surplus is the total utility (*TU*) you have gained from all the units minus your total expenditure (*TE*) on them: $TCS = TU - TE$. It is the sum of the marginal consumer surpluses from each unit you have purchased.

We can illustrate these ideas graphically. Figure 4.1 shows the marginal utility curve for a particular person and a particular good. If the price of the good were P_1, the person would consume Q_1, where $MU = P_1$. Thus point *a* would be one point on that person's demand curve. If the price fell to P_2, consumption would rise to Q_2, since this is where $MU = P_2$. Thus point *b* is a second point on the demand curve. Likewise if price fell to P_3, Q_3 would be consumed. Point *c* is a third point on the demand curve.

Because rational individuals consume where $MU = P$, their demand curve will be along the same line as their marginal utility curve.

The market demand curve

The market demand curve is simply the *horizontal* sum of all individuals' demand curves and hence *MU* curves.

The shape of the demand curve. The price elasticity of demand will reflect the rate at which *MU* diminishes. If there are close substitutes for a good then its *MU* will diminish slowly as its consumption increases. The reason is that increased consumption of this product will be accompanied by decreased consumption of the alternative products. People will have switched from alternatives and so substituted this product for those alternatives. Since total consumption of this product *plus* the alternatives has increased only slightly (if at all), the marginal utility will fall only slowly.

TC 10 p 49

KI 7 p 24

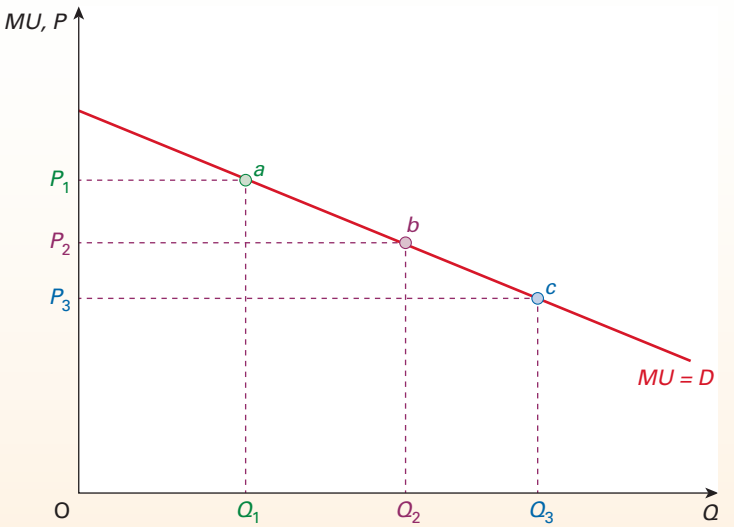
Shifts in the demand curve. A change in the consumption of one good affects the marginal utility of substitute and complementary goods. Hence, if the price of an *alternative* product in consumption were to change, then the *MU* schedule for a

Definitions

Marginal consumer surplus The excess of utility from the consumption of one more unit of a good (*MU*) over the price paid: $MCS = MU - P$.

Total consumer surplus The excess of a person's total utility from the consumption of a good (*TU*) over the total amount that person spends on it (*TE*): $TCS = TU - TE$.

Figure 4.1 Deriving an individual's demand curve



particular product will shift. For example, consider the marginal utility of (and hence demand for) margarine following a rise in the price of butter. The higher price of butter would cause less butter to be consumed. This would increase the marginal utility of margarine, since if people are using less butter, their desire for margarine is higher. The *MU* curve (and hence the demand curve) for margarine thus shifts to the right.

KI 7 p 24

Pause for thought

How would marginal utility and market demand be affected by a rise in the price of a complementary good?

Problems with simple marginal utility analysis

Deriving a demand curve from a marginal utility curve measured in money assumes that money itself has a constant marginal utility. The trouble is that it does not. If people have a rise in income, they will consume more. Other things being equal, the marginal utility of the goods that they consume will diminish. Thus an extra £1 of consumption will bring less satisfaction than previously. In other words, it is likely that the *marginal utility of money diminishes as income rises*.

Unless a good occupies only a tiny fraction of people's expenditure, a fall in its price will mean that their real income has increased: i.e. they can afford to purchase more goods in general. As they do so, the marginal utility of their money will fall.

We cannot, therefore, legitimately use money to measure utility in an absolute sense. We can, however, still talk about the relative utility that we get from different combinations of goods and show how we can make rational choices between them. We do this next.

The rational consumer's optimal combination of goods

A more realistic and satisfactory explanation of consumer choice is when we consider how individuals choose to allocate their scarce income across a range of goods. Then the analysis becomes one of determining a consumer's optimal combination of goods.

TC 2 p 9

The rule for rational consumer behaviour is known as the **equi-marginal principle in consumption**. This states that a consumer will achieve the highest utility from a given level of income when the ratio of the marginal utilities of goods is equal to the ratio of their prices.

For example, suppose that there are two goods: good A and good B. Assume that the last unit of good A you consumed gave three times as much additional utility as the last unit of B and yet good A only cost twice as much as good B. You would obviously gain by increasing your consumption of A and cutting your purchases of B. But as you switched from B to A, the marginal utility of A would fall due to diminishing marginal utility, and conversely the marginal utility of B would rise.

To maximise utility you would continue this substitution of A for B until the ratios of the marginal utilities (MU_A/MU_B) equalled the ratio of the prices of the two goods (P_A/P_B). Box 4.2 explains further the equi-marginal principle in consumption.

The **equi-marginal principle** can in fact be applied to all rational choices between alternatives, whether in consumption, as in this case, in production, employment or whatever. The principle follows from our understanding of rational decision making: people weigh up the marginal cost and marginal benefit of any activity. If the ratio of the marginal benefit of activity A to the marginal benefit of B were greater than the ratio of the marginal cost of A to B:

TC 2 p 9

$$\frac{MB_A}{MB_B} > \frac{MC_A}{MC_B}$$

then it would be rational to do more of A and less of B. You would carry on doing more of A and less of B until the two became equal.

$$\frac{MB_A}{MB_B} = \frac{MC_A}{MC_B}$$

At that point your total net benefit would be maximised.

When allocating scarce resources between alternatives, therefore, rational behaviour results in the marginal benefit ratios of the two alternatives equalling their marginal cost ratios.

Pause for thought

How do the concepts of marginal benefit and marginal cost ratios apply to rational consumption choices?

Definitions

Equi-marginal principle (in consumption) Consumers will maximise total utility from their incomes by consuming that combination of goods where the ratio of the marginal utilities of goods is equal to the ratio of their prices.

Equi-marginal principle (in general). The optimum amount of two alternatives consumed (or produced) will be where the marginal benefit ratios of the two alternatives are equal to their marginal cost ratios.

BOX 4.2 **OPTIMAL CONSUMPTION BUNDLES** EXPLORING ECONOMICS

Equi-marginal principle in consumption

TC 2 p 9

We can use marginal utility analysis to show how a rational person decides what combination of goods to buy. Given that we have limited incomes, we have to make choices. It is not just a question of choosing between two obvious substitutes, like a holiday in Greece and one in Spain, but about allocating our incomes between all the goods and services we might like to consume. If you have, say, an income of £20 000 per year, what is the optimum 'bundle' of goods and services for you to spend it on?

Making consumption choices at the margin

TC 2 p 9

The rule for rational consumer behaviour is known as the *equi-marginal principle*. This states that a consumer will get the highest utility from a given level of income when the ratio of the marginal utilities is equal to the ratio of the prices. Algebraically, this is when, for any pair of goods, A and B, that are consumed:

$$\frac{MU_A}{MU_B} = \frac{P_A}{P_B} \quad (1)$$

The left-hand side of equation (1) can be read as the relative size of the additional utility from the consumption last unit of A compared with that from the last unit of B. Hence, if MU_A/MU_B is 3, the additional utility from the last unit consumed of A is three times that from B. The right-hand side of the equation is the relative price of good A compared to that of good B. If P_A/P_B is also 3 then the price of A is three times that of B. In this case, there is no gain to be had by switching from one good to another.

Assume now that the price of good A falls so that P_A/P_B falls to 2, i.e. the price of good A is now twice that of good B. Since the last unit of good A consumed provided three times as much additional utility as the last unit of B, the rational consumer will buy more of good A and reduce consumption of good B. As more units of A and less of B are purchased, the marginal utility of A falls while that of B rises. The consumer will continue to buy more units of A and less of B until the last unit of A consumed adds twice as much utility to that from the last unit of good B.

KI 14 p 72

We can repeat the analysis for a rise in the price of good A. This time the consumer cuts backs on good A and increases consumption of good B. The consumer will do this up to the point that equation (1) is satisfied. This is the optimum combination of goods to consume.

An alternative way of presenting the equi-marginal principle is found by rewriting equation (1) as:

$$\frac{MU_A}{P_A} = \frac{MU_B}{P_B} \quad (2)$$

The reasoning is simple. If, for any consumer, good A had a higher marginal utility relative to its price than good B relative to its price:

$$\frac{MU_A}{P_A} > \frac{MU_B}{P_B}$$

that person should buy more A and/or less B. On doing so, the marginal utility of A will fall because of diminishing marginal utility and/or the marginal utility of B will rise until the MU/P ratios of the two goods were equal. In fact, we can generalise this principle to any number of goods. So in the case of n goods, the principle becomes:

$$\frac{MU_A}{P_A} = \frac{MU_B}{P_B} = \frac{MU_C}{P_C} \cdots = \frac{MU_n}{P_n} \quad (3)$$

The rational consumer's demand curve

We can use equation (3) to derive a demand curve for any good – say good A. We start with a given price of good A and the amount purchased that satisfies equation (3). This gives one point on the demand curve for good A.

Now we see what happens when we change the price of A. If the price of good A falls, the rational consumer would buy more of good A and less of other goods (B, C, D, E, etc.) until equation (3) is once more satisfied. A second point on the individual's demand curve for good A has been determined. (Note that the changes in the amounts consumed of the other goods will vary according to how close they are as substitutes or complements. The closer they are, the more their consumption will change as the price of A changes.)

Further changes in the price of good A would bring further changes in the quantity demanded, in order to satisfy equation (3). Further points on the individual's demand curve would thereby be derived.

If the price of *another* good changed, or if the marginal utility of any good changed (including good A), then again the quantity demanded of good A (and other goods) would change, until the equation is satisfied. These changes in demand will be represented by a *shift* in the demand curve for good A.

1. *In maximising utility, does it matter what units are used to measure utility?*
2. *How could the equi-marginal principle be applied to a firm's production choices?*

Make a list of the prices and amounts of products you purchase at a supermarket on an occasion when you buy a lot of things. Now consider how your purchases would have been different if the prices of three of the items had been 50 per cent higher. Would the changes you would have made be in accordance with equation (3) above? Explain why or why not.

Recap

1. The satisfaction people get from consuming a good is called 'utility'. Total utility is the satisfaction gained from the total consumption of a particular good over a given period of time. Marginal utility is the extra satisfaction gained from consuming one more unit of the good.
2. The marginal utility tends to fall the more that people consume. This is known as the *principle of diminishing marginal utility*.
3. The utility that people get from consuming a good (good A) will depend on the amount of other goods they consume. A change in the amount of other goods consumed, whether substitutes or complements, will shift the total and marginal utility curves for good A.
4. Rational consumers will attempt to maximise their consumer surplus. Consumer surplus is the excess of people's utility (measured in money terms) over their expenditure on the good. This will be maximised by purchasing at the point where the *MU* of a good is equal to its price.
5. In the simple case where the price and consumption of other goods are held constant, a person's *MU* curve will lie along the same line as that person's demand curve.
6. The market demand curve is the horizontal sum of the demand curves of all the individual consumers. The elasticity of the market demand curve will depend on the rate at which marginal utility diminishes as more is consumed. This in turn depends on the number and closeness of substitute goods. If there are close substitutes, people will readily switch to this good if its price falls, and thus marginal utility will fall only slowly. The demand will be elastic.
7. Measuring the marginal utility of a good in money avoids the problem of using some imaginary unit such as utils, but it assumes that money has a constant utility. In reality, the marginal utility of money is likely to decrease as income rises.
8. A more realistic and satisfactory way of analysing the demand for goods is to look at people's choices between goods. A consumer will maximise utility from a given income by consuming according to the 'equi-marginal principle'. This states that goods should be consumed in that combination which equates the *MU/P* ratio for each good.

4.2 THE TIMING OF COSTS AND BENEFITS

Analysing choices when benefits and costs occur over time

Intertemporal choices

The exact timing of the costs incurred and the benefits received vary between different types of consumption. This has implications for the way the rational choice model is applied. In some cases, all the costs and benefits of a decision are virtually instantaneous, with only a very small delay in time between them. For example, if you purchase a coffee the cost is immediate (unless you use a credit card) and the total pleasure from consuming the drink occurs shortly afterwards.

However, for a whole range of other consumption decisions all the costs and benefits are not instantaneous and occur over a more prolonged period of time. There may also be significant delays between the point in time the costs are incurred and the benefits received. Choice made under these conditions is known as **intertemporal choice**. Choices made at one moment in time can therefore affect future options.

Take the example of buying a consumer durable, such as a mobile phone, dishwasher or car. The major cost of purchasing many of these products is often virtually immediate (unless paying by instalments), while the stream of benefits they provide occurs for months or years after the initial costs are paid.

In other cases, all the benefits from consumption are instantaneous, while some of the costs occur in the future. For example, the consumption benefits and monetary costs of purchasing alcohol, cigarettes or unhealthy food are fairly immediate, while the health costs occur in the future. For a rational person, it would be necessary to weigh up all these costs and benefits. In other words, intertemporal decision making would be required in each of these examples.

TC 2 p 9

Maximising utility with intertemporal choices

How can the rational choice model be extended to analyse and explain these types of intertemporal choices?

The standard economic theory of consumer choice assumes that most people are impatient most of the

TC 4 p 13

Definition

Intertemporal choices Choices where the costs and/or benefits occur over several periods and where therefore choices made at one moment in time affect future options.

time. They would prefer to consume the things they like immediately rather than having to wait until a later date. They would also prefer to delay any costs until later – for example, paying for goods using a credit card.

To capture this impatience, standard economic theory uses a method of weighting future costs and benefits. It is called **exponential discounting** and multiplies any costs and benefits that occur in the future by a fraction of less than one to adjust them to what they are worth to the person immediately: i.e. their **present value**. This fraction is called the **discount factor**. For example, if a person's impatience could be captured by a monthly discount factor of, say, 0.9, then £100 of consumption benefits to be received in a month's time would be valued at £90 (0.9 × £100) today. In other words, the person would only be prepared to pay £90 today to be guaranteed to receive £100 in a month's time.

Levels of impatience will vary from one individual to another. The more impatient people are, the lower their discount factor and the less they will value benefits and costs that occur at some point in the future. For example, £100 of consumption benefits in a month for an individual with a monthly discount factor of 0.4 would only be worth £40 (0.4 × £100) today.

Box 4.3 considers further intertemporal decision making and the measurement of impatience.

Pause for thought

What discount factor is used to weight benefits that occur in three months' time for a person with a per-monthly discount factor of 0.9?

4.3 UNCERTAINTY AND RISK

What to do when you can't be certain of outcomes

When analysing demand and supply decisions, it is important to recognise that people often have *imperfect information* when making choices.

Sometimes we do have perfect or near perfect information. For example, it is reasonable to assume that when we buy goods and services, we know exactly what price we will pay and how much we will enjoy it. When you buy a bar of chocolate, for example, you clearly do know how much you are paying for it and have a very good idea how much you will like it.

But what about a mobile phone, or a tablet, or a car, or a washing machine, or any other consumer durable? In each of these cases you are buying something that will last you a long time, and the further into the future you look, the less certain you will be of its costs and benefits to you. When you buy it, therefore, you are taking a bit of a gamble.

Consider now the problem facing buyers and sellers over future prices. They will try to anticipate price changes, but unfortunately on many occasions they cannot be certain just what these price changes will be.

Take the case of stocks and shares. If you anticipate that the price of, say, HSBC shares is likely to go up substantially in the near future, you may well decide to buy some now and then sell them later after the price has risen. But you cannot be *certain* that they will go up in price; they may fall instead. Again, therefore, decision making involves taking a gamble.

Choice under risk and uncertainty

So how does a lack of certainty affect people's behaviour and the choices they make? The answer is that it depends on their attitudes towards taking a gamble. To examine these attitudes let us assume that a person does at least know their *chances* when taking a gamble (i.e. the *probabilities involved* in doing so). In other words, the person is operating under conditions of **risk** rather than **uncertainty**. Under uncertainty, *the probability of an outcome is not known*.

Definitions

Exponential discounting A method of reducing future benefits and costs to a present value. The discount rate depends on just how much less, from the consumer's perspective, *future* utility and costs (from a decision made today) are than gaining the utility/incurring the costs *today*.

Present value (in consumption) The value a person places today on a good that will not be consumed until some point in the future.

Discount factor The value today of deciding to consume a good one period in the future as a proportion of the value when it is actually consumed.

Risk When an outcome may or may not occur, but its probability of occurring is known. It is a measure of the variability of that outcome.

Uncertainty When an outcome may or may not occur and its probability of occurring is not known.

BOX 4.3 **INTERTEMPORAL DECISION MAKING AND THE RATIONAL CONSUMER** **EXPLORING ECONOMICS**

Incorporating impatience into models of consumer choice

TC 2 p 9

For many consumption choices the costs and benefits can occur over a period of time and there may be a delay between the points at which the costs and benefits occur. For example, investment decisions, such as those relating to financial products or house purchase, have payoffs (cost and benefits) that occur over a period of time. Such choices are known as intertemporal choices.

TC 4 p 13

How can our model of the rational consumer be extended to analyse and explain these types of intertemporal choices?

Impatience and utility

Standard economic theory assumes people prefer to consume the things they like immediately rather than having to wait until a later date. This is the idea of impatience. Similarly, they also prefer to delay any costs until later.

Impatience can be illustrated by the following simple example. Imagine that it is 10am on Monday morning and you are given the choice between receiving a payment of £500 immediately or having to wait until 10am on Tuesday morning. Given this choice, most people would prefer to have the £500 immediately, even if it were absolutely guaranteed that, if they chose Tuesday, they would receive the money then.

The key to understanding this impatience is to think about the decision at the time it is being made. If a person prefers to receive £500 on Monday rather than having to wait until Tuesday, then the following must be true.

$$U^{\text{Monday}}\text{: } u(£500_{\text{Monday}}) > u(£500_{\text{Tuesday}})$$

This states that from the person's perspective on Monday (U^{Monday}), the utility from receiving the money on Monday ($u(£500_{\text{Monday}})$) is greater than the utility of having to wait until Tuesday ($u(£500_{\text{Tuesday}})$).

This does not mean that from their perspective on *Tuesday* that £500 received on Tuesday would give them any less pleasure that it would on Monday. In other words:

$$U^{\text{Monday}}\text{: } u(£500_{\text{Monday}}) = U^{\text{Tuesday}}\text{: } u(£500_{\text{Tuesday}})$$

From the person's perspective on Tuesday (U^{Tuesday}), £500 received on Tuesday provides the same utility as receiving the £500 on Monday from their point of view on Monday. Impatience in this example means that judging the decision from Monday's perspective, receiving £500 immediately would give the person more pleasure than having to wait 24 hours.

Measuring impatience

From Monday's perspective how much more utility does receiving £500 immediately provide rather than having to wait until Tuesday? To capture this impatience, standard economic theory uses an approach called *exponential discounting*. This weights future costs and benefits by a fraction of less than 1. This fraction is called the *discount factor* and it adjusts costs and benefits that occur in the future to what they are worth to the person immediately: i.e. their *present value*.

To illustrate this idea, assume a person is considering a consumption decision when all the costs occur immediately, while all the benefits occur at one point in time in the future. The good costs £25, which is payable immediately, and provides £30 of utility in exactly one month's time (perhaps there is a month's delay before it can be delivered). As all the costs are immediate, they are weighted at 100 per cent of their value. However, impatience means that the future benefits are weighted by a discount factor.

If the person's impatience could be captured by a monthly discount factor of, say, 0.9 then £30 of pleasure in a month's time would be worth £27(0.9 × £30) to that person today. The present value of the benefit from consuming the good (£27) would still be greater than the immediate cost (£25). Hence, a rational person would purchase the product and be prepared to wait for delivery.

Consider now a more impatient individual whose level of impatience can be captured by a discount factor of 0.8. This makes £30 of benefit received in a month's time worth only £24 to that person now. Since the immediate cost of £25 is now greater than the discounted value of the future benefit, this more impatient individual would not purchase the good.

The further into the future any costs and benefits occur, the greater their values have to be reduced to adjust them to what they are worth to an individual today. Assume, in our example, that the delay to delivery is now two months and that the discount factor is a constant 0.9 per month. Hence, the discount factor applied to the future consumption benefit is 0.81 (i.e. 0.9 × 0.9). A benefit of £30 in two months' time has a present value of £24.30 (i.e. 0.81 × £30). This is less than the immediate cost (£25) and so our rational individual will not now purchase the product and wait for delivery.

1. a. *What discount factor is used to weight benefits that occur in three months' time for a person with a discount factor of 0.9 per month?*
   b. *What does this make the present value of £20 of benefits received in three months' time?*
2. *Assume that the good costs £10, which has to be paid today. How long would the maximum delay in months before receiving the £20 of benefits have to be before a person with a monthly discount of 0.9 would no longer purchase the good?*

Undertake a literature search to identify the types of intertemporal choices analysed by economists. Try then grouping them by the branch or field of economics.

Imagine that as a student you only have £105 left out of your student loan to spend. You are thinking of buying an instant lottery ticket/scratch card. The lottery ticket costs £5 and there is a 1 in 10 (i.e. 10 per cent) chance that it will be a winning ticket. A winning ticket pays a prize of £50. Would you buy the lottery ticket? This will depend on your attitude towards risk.

In order to explain people's attitude towards risk it is important to understand the concept of expected value. The **expected value** of a gamble is the amount the person would earn on average if the gamble were repeated many times. To calculate the expected value of a gamble you simply multiply each possible outcome by the probability that it will occur. These values are then added together.

In this example the gamble has only two possible outcomes – you purchase a winning ticket or a losing ticket. There is a 10 per cent chance it is a winning ticket, which will give you a total of £150 to spend (£100 left out of your loan plus a £50 prize). There is a 90 per cent chance it is a losing ticket, in which case you will only have £100 left to spend out of your student loan. Therefore the expected value (EV) of this gamble is £105.

$$EV = 0.1(150) + 0.9(100) = 105$$

If you had not taken the gamble, you would have £105 to spend for sure.

Attitudes towards risk

We can identify three possible categories of attitude towards risk.

Risk neutral. If people are risk neutral, they will always choose the option with the highest expected value. Therefore, a student who is risk neutral would be indifferent between buying or not buying the instant lottery ticket, as each outcome has the same expected value of £105.

Risk averse. If people are risk averse, they will never choose a gamble if it has the same expected value as the pay-off from not taking a gamble. Therefore, a student who is risk averse would definitely not buy the instant lottery ticket.

It is too simplistic, however, to say that a risk averse person will never take risks. Such a person *may* choose a gamble if it has a greater expected value than the pay-off from not taking the gamble. If the probability of purchasing a winning instant lottery ticket in the previous example was 20 per cent instead of 10 per cent, then a risk averse student might buy the ticket, as the expected value of the gamble (£110) is greater than the certain payoff (£105).

Whether or not risk-averse people do take a gamble depends on the *strength* of their aversion to risk, which will vary from one individual to another. The greater people's level of risk aversion, the greater the expected value of a gamble they are willing to give up in order to have a certain payoff.

Risk loving. If people are risk loving, they would always choose a gamble if it had the same expected value as the payoff from not taking the gamble. Therefore, a risk-loving student would definitely purchase the instant lottery ticket.

Once again, it is too simplistic to say that risk-loving people will always choose a gamble. For example, if the probability of purchasing a winning instant lottery ticket in the previous example was 1 per cent instead of 10 per cent, then even a risk-loving student might choose not to buy the ticket. It would depend on the extent to which that person enjoyed taking risks. The more risk-loving people are, the greater the return from a certain payoff they are willing to sacrifice in order to take a gamble.

Pause for thought

If people are generally risk averse, why do so many people around the world take part in national lotteries?

KEY IDEA 15

People's actions are influenced by their attitudes towards risk. **Many decisions are taken under conditions of risk or uncertainty. Generally, the lower the probability of (or the more uncertain) the desired outcome of an action, the less likely will people be to undertake the action.**

Diminishing marginal utility of income and attitudes towards risk taking

Most people, for most of the time, are risk averse. We prefer to avoid insecurity. This is because risk-averse behaviour accords with the principle of KI 14 p 72

Definitions

Expected value The predicted or average value of an outcome over a number of occurrences, calculated by taking each of the possible outcomes and multiplying it by its probability of occurrence and then adding each of these values.

Risk neutral When a person is indifferent between a certain outcome and a gamble with the same expected value.

Risk averse Where you would require a gamble to have a higher expected value than a certain outcome before being willing to take the gamble. The more risk averse you are, the higher the expected value you would require (i.e. the better would have to be the odds).

Risk loving Where you would be willing to take a gamble even if its expected value was lower than that of certain outcome. The more risk loving you are, the lower the expected value you would be prepared to accept (i.e. the worse the odds would need to be).

diminishing marginal utility. In the case of each individual good, the more we consume, the less satisfaction we gain from each additional unit: the marginal utility falls. But the same principle applies if we look at our *total* consumption. The higher our level of total consumption, the less additional satisfaction will be gained from each additional £1 spent. What we are saying here is that there is a **diminishing marginal utility of income**. The more you earn, the lower is the utility gained from each *extra* £1.

Because of the diminishing marginal utility of income, the gain in utility to people from an extra £100 is less than the loss of utility from forgoing £100. Thus if people are offered the gamble of a 50:50 chance of winning or losing £100, they will probably decline the gamble. This means that risk aversion is part of rational utility-maximising behaviour.

Pause for thought

Does diminishing marginal utility of income provide a moral argument for redistributing income from the rich to the poor? Does it prove that income should be so redistributed?

Most of the time we do not know the exact chances involved of taking a gamble. In other words, we operate under conditions of *uncertainty*. We often have to make judgements about what we think are the different likelihoods of various outcomes occurring. There is evidence that in some circumstances people are not very good at making probabilistic judgements and are prone to making systematic errors. This is discussed in more detail in Section 4.4.

Insurance: a way of removing risks

Insurance is the opposite of gambling. It removes the risk. If, for example, you risk losing your job if you are injured, you can remove the risk of loss of income by taking out an appropriate insurance policy.

KI 15 p 80

Given that many people are risk averse, they may be prepared to pay the premiums even though it will leave them with less than the expected value from taking the gamble.

The total premiums paid to an insurance company will be *more* than the amount it pays out: that is how such companies make a profit. An insurance company is prepared to shoulder the risks that its customers are not prepared to because it is able to **spread (or pool) its risks**.

Because an insurance company will be insuring many customers, for example against property burning down, it will be able to collect more than enough to cover its payments. The more clients it insures, the smaller will be the variation in the proportion of claims it will need to pay out on each year. This is an application of the **law of large numbers**. In other words, the more people the insurance company insures, the more predictable is the total outcome. Hence, an insurance company is able to convert your *uncertainty* into their *risk*.

The spreading of risks does not just require a large number of policies; it also requires the risks to be **independent**. If an insurance company insured properties against fire damage and they were *all in the same neighbourhood,* and then there was a fire in the area, the claims would be enormous; the risks of fire were not independent. If, however, it provides fire insurance for houses scattered all over the country, the risks *are* independent.

Pause for thought

Explain why an insurance company could not pool the risk of flooding in a particular part of a country. Does your answer imply that insurance against flooding is unobtainable?

Another way in which insurance companies can spread their risks is by **diversification**. The more types of insurance a company offers (car, house, life, health, etc.), the greater the likelihood the risks will be independent.

Definitions

Diminishing marginal utility of income Where each additional unit of income earned yields less additional utility than the previous unit.

Spreading (or pooling) risks (for an insurance company) The more policies an insurance company issues and the more independent the risks from these policies are, the more predictable will be the number of claims.

Law of large numbers The larger the number of events of a particular type, the more predictable will be their average outcome.

Independent risks Where two risky events are unconnected. The occurrence of one will not affect the occurrence of the other.

Diversification Where a firm expands into new types of business.

BOX 4.4 FUTURES MARKETS

EXPLORING ECONOMICS

A way of reducing uncertainty

KI 15 p 80

When faced with uncertainty, people have to form judgements about what they think are the different likelihoods of various outcomes occurring. One way of reducing or even eliminating uncertainty is by dealing in *futures* or *forward markets*. Let us examine first the activities of sellers and then those of buyers.

Sellers

Suppose you are a wheat farmer and want to store grain to sell at some time in the future, expecting to get a better price then than now. The trouble is that there is a chance that the price will go down. Given this uncertainty, you may be unwilling to take a gamble.

An answer to your problem is provided by the *commodity futures market*. This is a market where prices are agreed between sellers and buyers today for delivery at some specified date in the future.

For example, if it is 20 October today, you could be quoted a price today for delivery in six months' time (i.e. on 20 April). This is known as the six-month ***future price***. Assume that the six-month future price of wheat is £160 per tonne. If you agree to this price and make a six-month forward contract, you are agreeing to sell a specified amount of wheat at £160 on 20 April. No matter what happens to the ***spot price*** (i.e. the current market price) in the meantime, your selling price has been agreed. The spot price could have fallen to £140 (or risen to £180) by April, but your selling price when 20 April arrives is fixed at £160. There is thus *no risk to you whatsoever of the price going down*. You will, of course, have lost out if the spot price is *more* than £160 in April.

Buyers

Now suppose that you are a flour miller. In order to plan your expenditures, you would like to know the price you will have to pay for wheat, not just today, but also at various future dates. In other words, if you want to take delivery of wheat at some time in the future, you would like a price quoted *now*. You would like the risks removed of prices going *up*.

Let us assume that today (20 October) you want to *buy* the same amount of wheat on 20 April that a farmer wishes to sell on that same date. If you agree to the £160 future price, a future contract can be made with the farmer. You are then guaranteed that purchase price, no matter what happens to the spot price in the meantime. There is thus *no risk to you whatsoever of the price going up*. You will, of course, have lost out if the spot price is *less* than £160 in April.

The determination of the future price

Prices in the futures market are determined in the same way as in other markets: by demand and supply. For example, the six-month wheat price or the three-month coffee price will be that which equates the demand for those futures with the supply. If the five-month sugar price is currently £250 per tonne and people expect by then, because of an anticipated good beet harvest, that the spot price for sugar will be £200 per tonne, there will be few who want to buy the futures at £250 (and many who want to sell). This excess of supply of futures over demand will push the five-month future price down.

Speculators

Many people operate in the futures market who never actually handle the commodities themselves. They are neither producers nor users of the commodities. They merely speculate. Such speculators may be individuals, but they are more likely to be financial institutions.

KI 13 p 61

Let us take a simple example. Suppose that the six-month (April) coffee price is £4400 per tonne and that you, as a speculator, believe that the spot price of coffee is likely to rise above that level between now (October) and six months' time. You thus decide to buy 20 tonnes of April coffee futures now.

But you have no intention of taking delivery. After four months, let us say, true to your prediction, the spot price (February) has risen and as a result the April price (and other future prices) have risen too. You thus decide to *sell* 20 tonnes of April (two-month) coffee futures, whose price, let us say, is £4600. You are now 'covered'.

When April comes, what happens? You have agreed to buy 20 tonnes of coffee at £4400 per tonne and to sell 20 tonnes of coffee at £4600 per tonne. All you do is to hand the futures contract to buy to the person to whom you agreed to sell. They sort out delivery between them and you make £200 per tonne profit.

If, however, your prediction had been wrong and the price had *fallen*, you would have made a loss. You would have been forced to sell coffee contracts at a lower price than you bought them.

Speculators in the futures market thus incur risks, unlike the sellers and buyers of the commodities, for whom the futures market eliminates risk. Financial institutions offering futures contracts will charge for the service: for taking on the risks.

If speculators believed that the price of cocoa in six months was going to be below the six-month future price quoted today, how would they act?

Undertake desktop research to identify the range of different types of futures markets.

Definitions

Future price A price agreed today at which an item (e.g. commodities) will be exchanged at some set date in the future.

Spot price The current market price.

Choices under asymmetric information

In many cases economic choices are being made where there is significant **asymmetric information**. Asymmetric information exists in a market if one party has some information that is relevant to the value of that transaction that the other party does not have. In the insurance market, the buyer often has private information about themselves (their health, their behaviour, etc.) that the insurance company does not have access to.

Asymmetric information is often split into two different types – unobservable characteristics and unobservable actions. Each separate type of asymmetric information can potentially generate a different problem. Unobservable characteristics could generate the problem of *adverse selection;* unobservable actions could generate the problem of *moral hazard*. We consider each in turn.

Unobservable characteristics and adverse selection

Unobservable characteristics are present in many markets and may relate to the buyer, the seller or the product that is being traded. For example, in the case of second-hand car markets the unobserved characteristic could be the quality of car being brought to the market, while in credit markets it could be the ability of the borrower to manage their finances. The result is that there is an *informed* party (the individual selling a car or the borrower) and a *uniformed* party (a car salesperson/customer or a financial institution).

Unobservable characteristics can generate the problem of **adverse selection**. This is a market process whereby either buyers, sellers or products with certain unobservable adverse characteristics (e.g. high risk or low quality) are more likely to enter the market at the current market price. This process can have a negative impact on economic efficiency and cause some potentially profitable markets to collapse.

In the second-hand car market, for example, if dealers offered to buy cars at the average worth of low-quality cars ('lemons') and high-quality cars ('peaches'), the market could see largely low-quality cars being brought to sale. Similarly, if lenders in the credit market charged interest rates which incorporated an average risk premium across all borrowers, the market could become dominated by high-risk borrowers.

Unobservable actions and moral hazard

KI 15 p 80

Unobservable actions can generate the problem of **moral hazard**. This is where the actions or behaviour of one party to a transaction change once a deal has been reached. This then reduces the payoff to the other party.

It is caused by a change in incentives following the deal being reached. For example, if you hire a car, you may be rough with the clutch or gears, knowing that you will not bear the cost of the extra wear and tear of the car. A particularly salient example, following the financial crisis of the late 2000s, concerns bank bailouts. If a bank knows that it will be bailed out by the government and not allowed to fail, it may undertake more risky lending strategies (we will consider this argument further in Chapters 11 and 13). A similar example is the spending behaviour of young people. If they believe that parents will probably help them out if they get into debt, it is likely to make them less careful in their spending!

Box 4.5 considers the concepts of adverse selection and moral hazard in the context of insurance markets.

Tackling the problems of adverse selection and moral hazard

Tackling the problem of adverse selection. The party who is uninformed about the relevant characteristics could ask the other party for information. For example, an insurance company may require people to fill out a questionnaire giving details about their lifestyle and family history, or undergo a medical, so that the company can better assess risk and set an appropriate premium (see Box 4.5). This may need to be backed up by legal penalties for people caught lying. This process of the uninformed trying to get the information from the informed is called *screening*.

An alternative would be for the person or party who is informed about the relevant characteristics taking action to reveal it to the uninformed person or party. This is called *signalling*. For example, an intelligent and potentially hardworking person looking for a job could signal this fact to potential employers by obtaining a good grade in an economics degree or working for a period of time as an unpaid intern. This would encourage the firm to employ them rather than someone else.

Definitions

Asymmetric information Where one party in an economic relationship has more or better information than another.

Adverse selection A market process whereby either buyers, sellers or products with certain unobservable characteristics (e.g. high risk or low quality) are more likely to enter the market at the current market price.

Moral hazard The temptation to take more risk when you know that other people (e.g. insurers) will cover the risks.

Tackling moral hazard. What are the most effective ways of reducing moral hazard? One approach would be for the uninformed party to devote more resources to *monitoring* the actions and behaviour of the informed party to reduce the asymmetry of information. Examples include: insurance companies employing loss adjusters to assess the legitimacy of claims; lecturers using plagiarism detection software to discourage students from attempting to pass off other people's work as their own. However, monitoring can often be difficult and expensive.

An alternative is to change the terms of the deal so that the party with the unobservable actions has an incentive to behave in ways which are in the interests of the uninformed party. Examples include: the use of excesses in insurance, where the insured has to pay the first so much of any claim; employees who take sick leave being required to produce a medical certificate to prevent people taking 'sickies'; students doing group project work being assessed on their own contribution to the project rather than being given the same mark as everyone else in the group, thereby discouraging free riding.

Pause for thought

How can potential employees signal their ability to a potential employer? How could employers screen potential employees?

BOX 4.5 PROBLEMS WITH INSURANCE MARKETS EXPLORING ECONOMICS

Adverse selection and moral hazard

Two problems encountered by insurance companies in setting insurance premiums (the price of insurance) are termed *adverse selection* and *moral hazard*. Both these problems arise as a consequence of *asymmetric information,* when the insurance company knows less about the individuals seeking insurance than do the people themselves.

Adverse selection

This is where the people who take out insurance are likely to be those who have the highest risk.

For example, suppose that a company offers health-care insurance. It surveys the population and works out that the average person requires £200 of treatment per year. The company thus sets the premium at £250 (the extra £50 to cover its costs and provide a profit). But it is probable that the people most likely to take out the insurance are those most likely to fall sick: those who have been ill before, those whose families have a history of illness, those in jobs that are hazardous to health, etc. These people on average may require £500 of treatment per year. The insurance company would soon make a loss.

But could not the company then simply raise premiums to £550 or £600? It could, but it would thereby be depriving the person of average health of reasonably priced insurance and reducing the potential to sell insurance to people who value it.

The answer is for the company to obtain information that will allow it to identify the risk different people face. This process is known as *screening*. There are several methods that the company can adopt. It can ask for personal information. For example, if you are applying for health-care insurance, you may have to fill out a questionnaire giving details of your lifestyle and family history or undergo a medical so that the company can have an expert opinion on your level of risk and set an appropriate premium.

The second form of screening occurs when an insurance company makes use of market data to assess risk. For example, companies offering life assurance can make use of information on life expectancy.

Moral hazard

This occurs when having insurance makes you less careful and thus increases your risk to the company. For example, if your bicycle is insured against theft, you may be less concerned to go through the hassle of chaining it up each time you leave it. Again, this is an example of asymmetric information because the insurance company cannot determine for any individual whether taking out insurance will make them more careless.

If insurance companies work out risks by looking at the total number of bicycle thefts, these figures will understate the risks to the company because they will include thefts from *uninsured* people who are likely to be more careful. One solution is for insurance companies to write contracts that allow them to reduce the incidence of moral hazard. For example, they could require the insured person to pay the first so much of any claim (an 'excess').

If the problems of adverse selection and moral hazard cannot be overcome by screening and carefully written contracts then this is not a problem for insurance companies alone. Most people in society are risk averse and are willing to pay others to reduce the risk that they face. Insurance markets have grown up in response to this. If insurance markets fail, then society suffers as a consequence.

KI 15 p 80

1. *What details does an insurance company require to know before it will insure a person to drive a car?*
2. *How will the following reduce moral hazard?*
   a. *A no-claims bonus.*
   b. *An excess, which means that the claimant has to pay the first part of any claim.*
   c. *Offering lower premiums to those less likely to claim (e.g. lower house contents premiums for those with burglar alarms).*

First, research the various types of insurance products available. Second, identify possible adverse selection and moral hazard issues arising with these various types of insurance and the ways in which insurers attempt to address them.

Recap

1. Many economic decisions are taken under conditions of risk or uncertainty. If we know the probability of an outcome occurring, we are said to be operating under conditions of *risk*. If we do not know the probability, we are said to be operating under conditions of *uncertainty*.
2. We can identify three different attitudes toward risk: risk loving, risk averse and risk neutral. Because of the diminishing marginal utility of income, it is rational for people to be risk averse (unless gambling is itself pleasurable).
3. Insurance markets provide a way of eliminating risk. If people are risk averse, they are prepared to pay premiums to obtain insurance. Insurance companies are able to pool risk by selling a large number of policies, but require that the risks are independent.
4. Both adverse selection and moral hazard are likely to occur in a range of economic relationships whenever there is a problem of asymmetric information. They can both be reduced by tackling the asymmetry. Better information can be provided by screening, signalling or monitoring, or there can be incentives for providing more accurate information.

4.4 BEHAVIOURAL ECONOMICS

Do people always behave 'rationally'?

What is behavioural economics?

The field of behavioural economics has developed rapidly over recent years. It integrates ideas from disciplines such as psychology, neuroscience and sociology into standard economic theory. These disciplines provide insights into human behaviour and human interaction. The development of modern behavioural economics is not an attempt to replace mainstream economic theory but rather to complement and enhance existing theory.

Here we focus on behavioural insights in the context of consumer behaviour. In Chapter 6 we look at the contribution of behavioural economics to explaining the behaviour of firms.

Traditional theory and behavioural economics

To understand further what behavioural economics is, it is useful to think back to the traditional economic approach. This assumes that people make rational choices. However, this does not mean that economists actually believe that everyone in the real world does so. They accept that real human beings make mistakes, behave impulsively and often care about the payoffs to others.

TC 2 p 9

What economists are assuming is that the people in their theoretical *models* behave in a rational and selfish manner. This is an example of *abstraction*. If theories built upon this simplified view of human behaviour can effectively explain and predict real-world behaviour, then it is a useful assumption to make.

The alternative would be to assume that people in the theories are as complicated as their real-word counterparts. This would introduce much greater complexity into the analysis and make it much more difficult to understand and apply. If the simpler model is doing a good job at explaining and predicting real-world behaviour, why make it any more complicated than it needs to be?

Economists also argue that rational choice theory generally approximates human behaviour because, while people will tend to make mistakes, they do so in a haphazard and random manner that cancel each other out. Furthermore, people will learn how to make rational decisions from experience.

In many instances standard economic theory does a pretty good job, which explains why this textbook and many others are full of economic theories built on these assumptions. But sometimes, people's behaviour seems to run counter to traditional theory. We examine such behaviour in the remainder of the chapter.

Processing limited information

One reason why consumers do not always appear to behave rationally is that they have limited information. Choices are too complex, or the features of products are not fully understood or the outcomes are too uncertain for people to be able to maximise their consumer surplus.

Bounded rationality

Despite limited information and the complexity of choices, people may still want to maximise the utility they gain from their income. But, although on many occasions it *would* be possible to obtain better information, people decide it is not worth the time and effort, and perhaps expense, of doing so. Their ability to be 'rational' is thus limited or **bounded** by the situation in which they find themselves. Behavioural economists seek to understand the different assumptions people make and their different responses in situations of bounded rationality.

Sometimes the problem facing consumers is too much choice. They do not have the time and patience to look through all the alternatives.

Sometimes the problem is limited information, which means people may not know the exact costs and benefits of many decisions for certain. Limited information is a particular problem with consumer durables. For example, when purchasing a laptop computer, a customer may be unsure about (a) how well it will perform, (b) how useful it will be and (c) the chances it will break down.

To simplify the choice problem, people often revert to using mental short-cuts, referred to as **heuristics**, which require only modest amounts of time and effort. A heuristic technique is any approach to problem-solving, such as deciding what to buy, which is practical and sufficient for the purpose, but not necessarily optimal.

For example, people may resort to making the best guess, or to drawing on past experiences of similar choices that turned out to be good or bad. Sometimes, when people are likely to face similar choices again, they resort to *trial and error*. They try a product. If they like it, they buy it again; if not, they don't.

On other occasions, they may use various *rules of thumb*: buying what their friends do or buying products on offer or buying trusted brands. These rules of thumb can lead to estimates that are reasonably close to the utility people will actually get and can save on time and effort. However, they sometimes lead to systematic and predictable misjudgements about the likelihood of certain events occurring.

Framing and nudges

In traditional models of consumer choice, individuals aim to maximise their utility when choosing between goods, or bundles of goods. The context in which the choices are offered is not considered.

Definitions

Bounded rationality When the ability to make rational decisions is limited by lack of information or the time necessary to obtain such information.

Heuristics People's use of strategies that draw on simple lessons from past experience when they are faced with similar, although not identical, choices.

Framing The way in which a choice is presented or understood. A person may make different decisions depending on whether a choice is presented optimistically or pessimistically.

Nudge theory The theory that positive reinforcement or making the decision easy can persuade people to make a particular choice. They are 'nudged' into so doing.

Reference dependence The situation where people evaluate the outcomes of choices relative to a reference point and then classify them as either gains and losses.

Yet, in real life, we see that context is important; people will often make different choices when they are presented, or **framed**, in different ways. For example, people will buy more of a good when it is flagged up as a special offer than they would if there is no mention of an offer, even though the price is the same.

This principle has led to the development of **nudge theory**, which underpins many marketing techniques. Here people can be persuaded to make a particular choice by framing it in an optimistic way or presenting it in a way that makes it easy to decide. We look at governments' use of nudge theory in Box 4.6.

Reference points

Reference dependence. Where choices reflect risk or uncertainty, framing can also affect decision making. One reason for this is that people evaluate choices by forming reference points. They then class outcomes as gains and losses. Hence, decisions can be affected by **reference dependence**.

To illustrate reference dependence, consider the following two ways in which an equivalent choice is framed.

- 'If you switch to using this environmentally friendly product it will save you £200 per year in lower energy costs.'
- 'If you don't switch to using this environmentally friendly product you are wasting £200 per year in higher energy costs.'

The first presentation may lead to people perceiving the £200 as a potential *gain* since the reference point is the current situation before the choice is made. The second presentation moves the consumer's reference point to after the decision to switch or not. Then, the £200 is perceived as a *loss* of not switching. If people are more sensitive to a potential loss than a forgone gain, then presenting the choice in the second way is more likely to persuade people to purchase the environmentally friendly product.

Pause for thought

Assume that Dean and John are both students on an economic course. They have both exerted the same level of effort writing an assessed essay and each receive a mark of 60 per cent. Will they both be equally happy?

Reference dependent loss aversion. Because of the diminishing marginal utility of income (see page 81), people are generally loss averse. The additional utility from receiving a given amount of money is less than the loss of utility if they were to forgo the same amount.

KI 14 p 72

However, this loss aversion can be amplified by the reference point for a decision. This gives rise to the concept of **reference dependent loss aversion**.

KI 15 p 80

One way in which loss aversion might be greater than predicted by diminishing marginal utility of income in standard theory is when people are faced with disinvesting, selling or losing something they currently own. People tend to ascribe more value to things when they already own them, and are faced with selling them or otherwise giving them up, than when they are merely considering purchasing or acquiring them in the first place. In other words, when the reference point is one of ownership of an item, people put a higher value on it than when the reference point is one of non-ownership. This is known as the **endowment effect**, also referred to as **divestiture aversion**.

KI 15 p 80

The endowment effect may explain why people may stick with certain products, including financial investments and housing, even if unprofitable or no longer appropriate for need. Therefore, it can explain **status quo bias**: the tendency for people to prefer things to stay as they are or to continue with decisions already made.

Sunk cost fallacy. Loss aversion can help to explain a phenomenon known as the **sunk cost fallacy**. This is where people continue with their current behaviour or activities because of the resources, such as money, time and effort, previously invested. These are known as 'sunk costs' because the money or effort has already been invested and you will not get it back (see page 96). Sunk costs create emotions that cause people to feel committed to a particular course of action. Hence, sunk costs are another cause of status quo bias.

KI 15 p 80

An example of this arises if you find yourself reading a novel, but not enjoying it. What should you do? The answer is that, unless you believe that you will start enjoying the story soon, you should stop reading it. Many people would argue that they have paid for the book and so should carry on to the end. But this is not rational. It doesn't matter whether the book cost £5 or £50, the decision to read on should only take the future (opportunity) costs and future benefits into account.

Definitions

Reference dependent loss aversion Where people classify outcomes as either losses or gains in relation to a reference point and the perceived losses are disliked more than would be predicted by standard diminishing marginal utility.

Endowment effect (or **divestiture aversion**) The hypothesis that people ascribe more value to things when they own them than when they are merely considering purchasing or acquiring them.

Status quo bias Where, other things being equal, people are averse to change, preferring things to stay as they are.

Sunk cost fallacy The tendency for individuals to continue with activities because of the resources they have previously invested in them.

Taking other people into account

Our behaviour as consumers, as in many other aspects of our lives, is often influenced by other people – both the effect we have on them and the effect they have on us.

Relativity

If I am making a choice about buying a car, traditional economics says my demand will derive from a number of factors: my income; my tastes for driving and for particular cars; the prices of the car I am considering and of the alternatives; and the associated costs of motoring. Yet I might also be highly influenced by the car my sister drives; if she chooses an Audi, perhaps I would like a more expensive car – a Mercedes possibly. If she switches to a Jaguar, then perhaps I will opt for a Porsche. I want a better (or faster or more expensive) car than my sister; I am concerned not only with my choice of car but with my *relative* choice.

This does not disprove that our choices depend on our perceived utility. But it does demonstrate that our satisfaction often depends on our consumption *relative* to that of other people, such as our relatives and peers.

The act of comparing ourselves to others or 'keeping up with the Jones's' illustrates how people frequently make choices by drawing comparisons and looking at things relatively. This can distort choices between various products or activities since people are not looking at their characteristics in their own right; rather, they are looking at them in a particular context. This context *frames* the choice.

The importance of relativity in framing choices is something that retailers and advertisers are well aware of. For example, by showing customers the most expensive version of a product, customers may feel they have obtained a great deal by walking away with a slightly less expensive version, despite perhaps spending more than they had budgeted for.

Pause for thought

Can relativity explain why advertisers frequently show images of other *people consuming and enjoying products?*

Herding and 'groupthink'

Being influenced by what other people buy, and thus making relative choices, can lead to herd behaviour. A fashion might catch on; people might grab an item

in a sale because other people seem to be grabbing it as well; people might buy a particular share on the stock market because other people are buying it.

Now part of this may simply be bounded rationality. Sometimes it may be a good rule of thumb to buy something that other people want, as they might know more about it than you do. But there is a danger in such behaviour: other people may also be buying it because other people are buying it, and this builds a momentum. Sales may soar and the price may be driven well above a level that reflects the utility that people will end up gaining.

Reciprocity

TC 4 p 13

Behavioural economists have tried to model how people's utility is affected when they care not only about their own pay-offs but those of other people too.

For example, having *altruistic* preferences in economics means that you might be willing in some circumstances to increase the pay-offs to another person or group of people at a personal cost to yourself: e.g. making donations to a charity or buying presents. Having *spiteful* preferences, on the other hand, means that you might be willing to *reduce* the pay-offs of another person or group of people at a cost to yourself.

Evidence from experiments suggests that individuals can display both altruistic and spiteful behaviour. They are often willing to increase the pay-offs to others at a personal cost to themselves in some situations while reducing the pay-offs to others at a personal cost to themselves in other situations.

To capture these ideas, behavioural economists have developed models of **reciprocity**. Some of these suggest that people may experience an increase in utility by being kind to people who they believe have been kind to them. Their utility may also increase if they are unkind to people who they believe have been unkind to them.

Biased behaviour

Biases when assessing the likelihood of uncertain events

People's choices can be affected by various biases or preconceptions they may have. This then affects how they interpret and use information in decision making. It can make choices less than optimal. Three examples are listed below.

Friends and relatives as an information source. People tend to overweight the single experience of a friend or relative and underweight the experience of a large number of consumers you have never met. For example, assume that you are thinking of buying a particular model of laptop and are trying to determine its reliability. Evidence from consumer surveys based on the experiences of many users may suggest that it is reliable. However, a friend who has recently purchased the same model has had to send it back to the manufacturer for repairs. By focusing on your friend's experience, you may overestimate the chances of it breaking down.

Assigning too much weight to the most recent information. People tend to place considerable weight on recent information that can be retrieved from memory and not enough weight to older information. For example, you may have bought a particular type of cheese in the past but have not done so for a while. You see it again in a supermarket, but because you cannot remember how much you liked it you stick to a different type you have consumed recently and know you like.

Gambler's or Monte Carlo fallacy. This is the false belief that past outcomes have an impact on the likelihood of the next outcome occurring when, in reality, they are independent of one another. For example, assume a coin is tossed four times and comes up heads each time. You may mistakenly believe that a tail is more likely on the next toss. Another example is investors holding on to company shares after recent price falls in the belief that they must now go up.

Present bias and self-control issues

Traditional economic theory assumes that if people plan to do something at some specific point in time in the future, they do indeed do so when the time arrives. This is referred to as **time consistency**. The only reason time-consistent people would change their mind is if new information came to light about the relative size of the costs and benefits of their decisions.

For example, assume in the morning that you plan to go to the gym at 5pm. At 4pm, however, you change your mind because you find out that your best friend is coming to see you at 5pm. This is still time-consistent behaviour, as the only reason you have changed your mind is because information has changed: the opportunity cost of going to the gym turns out to be greater than you had earlier thought.

In practice, people often exhibit *time-inconsistent* behaviour. This is likely when some of the costs or benefits occur before others. People may then change their minds. For example, it comes around to 5pm and time to set off for the gym, but instead you chose to watch TV or play a computer game and plan to go to the gym tomorrow.

Definitions

Reciprocity (in economics) Where people's behaviour is influenced by the effects it will have on others.

Time consistency Where a person's preferences remain the same over time. For example, it is time consistent if you plan to buy a book when your student loan arrives and then actually do so when it does.

Many people make New Year's resolutions; most do not stick to them! People are weak willed; people put things off.

Pause for thought

Give some other examples of decisions where people often change their mind with the passage of time once the costs or benefits become immediate.

Behavioural economists refer to this form of time inconsistency as **present bias**. This is where people put a greater weight on present benefits and/or costs than would be implied by a standard approach to discounting the future (see Box 4.3).

An implication of present bias is that people will put excess weight on the costs of doing things they don't like doing but believe are good for them (such as going to the gym or dieting); and excess weight on the benefits of doing things they want to do but believe are bad for them (such as eating lots of chocolates or going out rather than revising).

Present bias can therefore be used to help explain why many people have difficulty in sticking to commitments and exhibit problems of self-control. In such situations, people may need some type of commitment mechanism that restricts their future choices by increasing the cost of taking certain actions. For example, this could be arranging to go to the gym with friends on certain days or buying unhealthy food in small packages even when considerable savings are available when it is purchased in far larger quantities.

Pause for thought

Assume you have a present bias but are fully aware of the inconsistent nature of your preferences. What actions could you take to make sure you carry out your planned decisions?

Implications for economic policy

Governments, in designing policy, will normally attempt to change people's behaviour – both consumers and producers. They might want to encourage people to choose to work harder, to save more, to recycle rubbish, to use their cars less, to eat more healthily, and so on. If the policy is to be successful, it is vital for the policy measures to contain appropriate incentives: whether it be a tax rise, a grant or subsidy, a new law or regulation, an advertising campaign or direct help. TC 6 p 24

But whether the incentives are appropriate depends on how people will respond to them, and to know that, the policy makers will need to understand people's behaviour. This is where behavioural economics comes in. People might respond as rational maximisers; but they may not. It is thus important to understand how context affects behaviour and adjust policy incentives appropriately.

Definition

Present bias Time-inconsistent behaviour whereby people give greater weight to present payoffs relative to future ones than would be predicted by standard discounting techniques.

Recap

1. Traditional economics is based on the premise that consumers act rationally, weighing up the costs and benefits of the choices open to them. Behavioural economics acknowledges that real-world decisions do not always appear rational; it seeks to understand and explain what economic agents actually do.
2. People's ability to make rational decisions is bounded by limited information and time. Thus, people may resort to using heuristics, such as rules of thumb.
3. The choices they make may also depend on how these choices are framed – the way in which they are presented or are perceived. People can be nudged to frame choices differently.
4. Seemingly irrational behaviour may arise from the choice of reference point for decision taking. The reference point used by people to judge an outcome as a gain or a loss can be influenced by a range of factors, including their expectations and comparisons with others.
5. People who are loss averse may value things more highly when they own them than when they are considering buying them (the endowment effect) or when the costs or benefits are immediate. This reference dependent loss aversion may result in people giving additional weight to loss than would occur simply from the diminishing marginal utility of income.
6. People may appear to behave irrationally because of a biased use of information: putting undue weight on the experience of friends or on their own experiences, especially recent ones. It may be because of the belief that independent things are really connected (the gambler's fallacy).
7. Giving additional weight to immediate benefits or costs is called present bias and can lead to time-inconsistent behaviour, with people changing their mind and not acting in accordance with previous plans.

BOX 4.6 **NUDGING PEOPLE** EXPLORING ECONOMICS

How to change behaviour without taking away choice

One observation of behavioural economists is that people make many decisions out of habit. They use simple rules (heuristics) such as: 'I'll buy the more expensive item because it's bound to be better'; or 'I'll buy this item because it's on offer'; or 'I'm happy with Brand X, so why should I change brands?'; or 'Other people are buying this, so it must be worth having'.

TC 6 p 24

Given that people behave like this, how might they be persuaded to change their behaviour? Governments might want to know this. What policies will encourage people to stop smoking, or save energy, or take more exercise or eat more healthy food? Firms too will want to know how to sell more of their products or to motivate their workforce. Even parents might want to make use of behavioural economics.

Opting in versus opting out

An interesting example concerns 'opting in' versus 'opting out'. In some countries, with organ donor cards, or many company pension schemes or charitable giving, people have to opt in. In other words, they have to make the decision to take part. Many as a result do not, partly because they never seem to find the time to do so, even though they might quite like to. With the busy lives people lead, it's too easy to think, 'Yes, I'll do that some time', but never actually get around to doing it. Hence, people exhibit *present bias*.

With an 'opt-out' system, people are automatically signed up to the scheme, but can freely choose to opt out. Take the case of organ donation. In the UK, this used to be an opt-in system to join the register. But, although 80 per cent of the public supported organ donation, less than 50 per cent ever got around to joining. In May 2020, the default position changed. Now people are automatically signed up for organ donation. If they do not want to donate their organs, they must opt out of the register. Early evidence shows relatively few people opting out.

Another move from opting in to opting out concerns UK company pension arrangements. Since 2012, firms must automatically deduct pension contributions from employees' wages unless they choose to opt out of the scheme. When employees had to opt in, 49 per cent enrolled in a company pension. When they had to opt out, the figure increased to 86 per cent enrolment.

Another example is that of charitable giving. Some firms add a small charitable contribution to the price of their products (e.g. airline tickets or utility bills), unless people opt out. Most people do not opt out.

This type of policy can improve the welfare of those who make systematic mistakes (i.e. suffer from present bias) while imposing very limited harm on those who act in a time-consistent manner. If it is in the interests of someone to opt out of the scheme, they can easily do so. Policies such as these are an example of what behavioural economists call *soft paternalism*.

The Behavioural Insights Team

The UK Coalition government (2010–15) established the Behavioural Insights Team (BIT) (also unofficially known as the Nudge Unit) in the Cabinet Office in 2010. BIT was partially privatised in 2014 and is now equally owned by the UK Government, the innovation charity Nesta and the Teams' employees.

A major objective of this team is to use ideas from behavioural science to design policies that enable people to make better choices for themselves. Some of this work involves the use of behavioural economics. For example, BIT ran a trial with the energy supplier Powershop Australia to investigate the impact of changing the default option for the company's 'Curb Your Power' policy. People who sign up for this scheme receive text messages during peak demand periods that ask them to reduce their use of power. The study found that households who were automatically enrolled in the scheme used 13.8 per cent less power during peak demand events than those households that had to sign up.

BIT developed a process, now widely used by policy-makers, to encourage changes in behaviour. It has the acronym EAST. The principles of EAST are:

- Make it **E**asy for people to change their behaviour by requiring them to make minimum effort and by making the changes simple to understand and execute. This was the reasoning behind the adoption of automatic enrolment in pension schemes for employees of large companies, where people had to opt out if they did not want to participate. Participation rates rose from 61 to 83 per cent.
- Make it **A**ttractive for people to change their behaviour, or unattractive for them not to, by giving incentives or improving the design of a product or activity. 'When letters to non-payers of car tax included a picture of the offending vehicle, payment rates rose from 40 to 49 per cent.'
- Make it **S**ocial, by showing other people behaving in the desirable way or encouraging people to make commitments to others. When people were informed by HMRC that most people pay their taxes on time, payment rates increased by as much as 5 percentage points.
- Make it **T**imely by prompting people at a time when they are likely to be the most receptive and by considering the direct costs and benefits of acting at the relevant moment in time. For example, sending text messages just before a payment is due or just before a person has to submit a form increases the response rate.

For government, nudging people to behave in ways that accord with government objectives can be both low-cost and effective.

1. *How would you nudge members of a student household to be more economical in the use of electricity?*
2. *How could the government nudge people to stop dropping litter?*
3. *In the 2011 Budget, the then Chancellor George Osborne announced that charitable giving in wills would be exempt from inheritance tax. Do you think this will be an effective way of encouraging more charitable donations?*

Visit the website of the Behavioural Insights Team (BIT). Search for Policy Publications and then write a short summary of the suggested approaches or interventions contained within one of the publications relating to a policy area of interest to you. What behavioural ideas are the motivation for the team's recommendations?

QUESTIONS

1. If two houses had identical characteristics, except that one was near a noisy airport and the other was in a quiet location, and if the market price of the first house were £360 000 and that of the second £400 000, how would that help us to put a value on the characteristic of peace and quiet?
2. Is it reasonable to assume that people seek to equate the marginal utility/price (*MU/P*) ratios of the goods that they purchase, if (a) they have never heard of 'utility', let alone 'marginal utility'; (b) marginal utility cannot be measured in any absolute way?
3. Consider situations where you might consider swapping items with someone. Why are such situations relatively rare? Can you think of circumstances in which this might be more common?
4. Explain why the price of a good is no reflection of the total value that consumers put on it?
5. How would marginal utility and market demand be affected by a rise in the price of a complementary good?
6. Give some examples of decisions you have taken recently that were made under conditions of uncertainty. With hindsight do you think you made the right decisions?
7. A country's central bank (e.g. the Bank of England or the US Federal Reserve Bank) has a key role in ensuring the stability of the banking system. In many countries the central bank is prepared to bail out banks which find themselves in financial difficulties. Although this has the benefit of reducing the chance of banks going bankrupt and depositors losing their money, it can create a moral hazard. Explain why.
8. Discuss the EU ruling that gender may not be used to differentiate insurance premiums. Which insurance markets would be affected by outlawing age 'discrimination' in a similar manner? What would be the impact?
9. Euro NCAP carries out crash tests on new cars in order to assess the extent to which they are safer than the minimum required standard. The cars are given a percentage score in four different categories, including adult occupant protection and child occupant protection. An overall safety rating is then awarded. Based on the test results in 2022, the Lexus NX was judged to be one of the safest cars on the market. If you observed that these cars were *more* likely to be involved in traffic accidents, could this be an example of adverse selection or moral hazard? Explain.
10. How does economics predict rational consumers will treat spending on credit cards compared with spending cash? Do you think that there are likely to be differences in the way people spend by each method? If so, can you explain why?
11. Give some example of heuristics that you use. Why do you use them?
12. For what reasons may branded products be more expensive than supermarkets' own-brand equivalents? How can behavioural economics help to explain this?
13. Why do gyms encourage people to take out monthly or even annual membership rather than paying per visit?
14. Many European countries operate organ donor schemes, some with schemes requiring that potential donors opt in, others with a system of opting out, or presumed consent. Explain why a system of presumed consent is likely to result in much higher numbers of donors.

Pearson | MyLab | Economics

Log on to MyLab Economics and complete the studyplan exercises for this chapter to see how much you have learnt and where you need to revise most. Make sure you access all the supporting textbook resources, including the online workbook, newsblog, audio animations, guided solutions and ebook.

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

4.1 **Bentham and the philosophy of utilitarianism.** This looks at the historical and philosophical underpinning of the ideas of utility maximisation.

4.2 **Utility under attack.** This looks at the birth of indifference analysis (see Web Appendix 4.2), which was seen as a means of overcoming the shortcomings of marginal utility analysis.

4.3 **The marginal utility revolution: solving the diamonds-water paradox.** This shows how marginal utility analysis can help explain why water which is so essential to human life has such a low market value compared with diamonds which while relatively trivial have such a high market value.

4.4 **Is economics the study of selfish behaviour?** To what extent do we take other people's interests into account when we make economic choices?

4.5 **The characteristics approach to analysing consumer demand.** An overview of the theory that people demand goods not for their own sake, but for the characteristics they possess.

(Continued)

4.6 **Love and caring.** An illustration of how rational choice behaviour (see Web Appendix 4.2) can be extended to family behaviour.

4.7 **Experimental economics.** This case explores how with the rise of behavioural economics, the use of experiments has become increasingly popular, both for research and teaching purposes.

4.8 **Modelling present bias.** An illustration of time-inconsistent behaviour: the tendency for people to change their plans with the passage of time once some of the costs or benefits have been experienced.

WEB APPENDICES

4.1 **Marginal utility theory.** This develops the analysis of Section 4.1 and considers the relationship between marginal utility and demand, not just for an individual product but for the choice between products.

4.2 **Indifference analysis.** This examines the choices consumers make between products and shows how these choices are affected by the prices of the products.

4.3 **Risk and uncertainty and the diminishing marginal utility of income.** An extension of the analysis in Section 4.3.

Chapter 5

The supply decision

So far we have assumed that supply curves are generally upward sloping: that a higher price will encourage firms to supply more. But just how much will firms choose to supply at each price? It depends largely on the amount of profit they will make. If a firm can increase its profits by producing more, it will normally do so.

Profit is made by firms earning more from the sale of goods than they cost to produce. A firm's total profit is thus the difference between its total sales revenue (*TR*) and its total costs of production (*TC*). In order then to discover how a firm can maximise its profit or even get a sufficient level of profit, we must first consider what determines costs and revenue.

In Sections 5.1 and 5.2 we examine short-run and long-run costs respectively. Over the short run a firm will be limited in what inputs it can expand. For example, a manufacturing company might be able to use more raw materials, or possibly more labour, but it will not have time to open up another factory. Over the long run, however, a firm will have much more flexibility. It can, if it chooses, expand the whole scale of its operations.

In Section 5.3 we turn to the revenue side and see how a firm's revenue varies with output. Then in Section 5.4 we put revenue and cost together to see how profit is determined. We see how profit varies with output and how the point of maximum profit is found.

In the final section we consider why firms' behaviour may deviate from the traditional profit-maximising model. It may be that in some circumstances firms may have an alternative aim; or perhaps there may be several potentially conflicting aims held by different stakeholders in the firm. We finish by considering how behavioural economics can be applied to the theory of the firm. In what ways might managerial decision making deviate from that predicted by rationality?

After studying this chapter, you should be able to answer the following questions:

- What is the relationship between inputs and outputs in both the short and long run?
- How do costs vary with output and just what do we mean by 'costs'?
- What are meant by 'economies of scale' and what are the reasons for such economies?
- How does a firm's sales revenue vary with output?
- How do we measure profits?
- At what output will a firm maximise its profits? How much profit will it make at this output?
- Why might firms choose not to maximise profits? What other objectives might firms follow?
- What sorts of mental short-cuts or rules of thumb might firms employ in their decision making?

5.1 PRODUCTION AND COSTS: SHORT RUN

How do a firm's costs vary with output over the short term?

Chapter 4 went behind the demand curve. It saw how the 'rational' consumer weighs up the *benefits* (utility) of consuming various amounts of goods or combinations of goods against their *costs* (their price).

We now need to go behind the supply curve and find out just how the **rational producer** (or 'firm' as we call all producers) will behave. In this case, we shall be looking at the benefits and costs to the firm of producing various quantities of goods and using various alternative methods of production. We shall be asking:

- How much will be produced?
- What combination of inputs will be used?
- How much profit will be made?

Short-run and long-run changes in production

If a firm wants to increase production, it will take time to acquire a greater quantity of certain inputs. For example, a manufacturer can use more electricity by turning on switches, but it might take a long time to obtain and install more machines, and longer still to build a second or third factory.

KEY IDEA 16

Output depends on the amount of resources and how they are used. **Different amounts and combinations of inputs will lead to different amounts of output. If output is to be produced efficiently, inputs should be combined in the optimum proportions.**

If, then, the firm wants to increase output relatively quickly, it will be able to increase the quantity of only certain inputs. It can use more raw materials, more fuel, more tools and possibly more labour (by hiring extra workers or offering overtime to its existing workforce). But it will have to make do with its existing buildings and most of its machinery.

The distinction we are making here is between **fixed factors** and **variable factors**. A *fixed* factor is an input that cannot be increased within a given time period (e.g. buildings). A *variable* factor is one that can.

The distinction between fixed and variable factors allows us to distinguish between the short run and the long run.

The short run. The **short run** is a time period during which at least one factor of production is fixed. In the short run, then, output can be increased only by using more variable factors. For example, if a chain of coffee shops became more successful, it could serve more customers per day in its existing shops – assuming there was space. It could increase the quantity of coffee beans and milk it purchases. It may be able to hire more staff, depending on conditions in the local labour market, and purchase additional coffee machines if there was space to install them. However, in the short run it could not extend its existing shops or have new ones built. This would take more time.

The long run. The **long run** is a time period long enough for all inputs to be varied. Given long enough, a firm can build additional factories and install new plant and equipment; a coffee chain can have new shops built.

The actual length of the short run differs from firm to firm. It is not a fixed period of time. Thus if it takes a farmer a year to obtain new land, buildings and equipment, the short run is any time period up to a year and the long run is any time period longer than a year. On the other hand, if it takes a manufacturer of mobile phone handsets two years to get a new factory built, the short run is any period up to two years and the long run is any period longer than two years.

Pause for thought

How long is the short run for an airline? How might this depend on the state of the aerospace industry?

For the remainder of this section we will concentrate on *short-run* production and costs. We will look at the long run in Section 5.2.

Production in the short run: the law of diminishing returns

Production in the short run is subject to *diminishing returns*. You may well have heard of 'the law of

Definitions

Rational producer or firm A firm which weighs up the costs and benefits of alternative courses of action and then seeks to maximise its net benefit.

Fixed factor An input that cannot be increased in supply within a given time period.

Variable factor An input that *can* be increased in supply within a given time period.

Short run The period of time over which at least one factor is fixed in supply.

Long run The period of time long enough for *all* factors to be varied.

diminishing returns': it is one of the most famous of all 'laws' of economics. To illustrate how this law underlies short-run production let us take the simplest possible case where there are just two factors: one fixed and one variable.

Take the case of a farm. Assume that the fixed factor is land and the variable factor is labour. Since the land is fixed in supply, output per period of time can be increased only by increasing the number of workers employed. But imagine what would happen as more and more workers crowded on to a fixed area of land. The land cannot go on yielding more and more output indefinitely. After a point the additions to output from each extra worker will begin to diminish.

We can now state the **law of diminishing (marginal) returns**.

KEY IDEA 17

The law of diminishing 'marginal' returns. **When increasing amounts of a variable factor are used with a given amount of a fixed factor, there will come a point when each extra unit of the variable factor will produce less extra output than the previous unit.**

Box 5.1 illustrates the law of diminishing returns in relation to a small retailer. Box 5.2 shows how the law can have potentially dire implications for us and the future inhabitants of our planet. Meanwhile Case Study 5.3 on the student website looks at diminishing returns to the application of nitrogen fertiliser on farmland.

The relationship between inputs and output is explained in more detail in Web Appendix 5.1, which looks at the short-run 'production function'.

Measuring costs of production

We are now ready to look at short-run costs. First of all, we will need to define just what we mean by costs. The term is used differently by economists and accountants.

TC 4 p 13

When measuring costs, economists always use the concept of *opportunity cost*. Remember from Chapter 1 how we defined opportunity cost (see page 7). It is the cost of any activity measured in terms of the *sacrifice* made in doing it: in other words, the cost measured in terms of the opportunities forgone.

How do we apply this principle of opportunity cost to a firm? First we must discover what factors of production it is using. Then we must measure the sacrifice involved. To do this it is necessary to put factors into two categories.

Factors not owned by the firm: explicit costs

The opportunity cost of those factors not already owned by the firm is simply the price that the firm has to pay for them. Thus if the firm uses £100 worth of electricity, the opportunity cost is £100. The firm has sacrificed £100 that could have been spent on something else.

Definition

Law of diminishing (marginal) returns When one or more factors are held fixed, there will come a point beyond which the extra output from additional units of the variable factor will diminish.

BOX 5.1 DIMINISHING RETURNS IN THE BREAD SHOP CASE STUDIES & APPLICATIONS

Is the baker using his loaf?

Just up the road from where I (John) live is a bread shop. Like many others, I buy my bread there on a Saturday morning. Not surprisingly, Saturday morning is the busiest time of the week for the shop and as a result it takes on extra assistants.

During the week only one assistant serves the customers, but on a Saturday morning there used to be five serving. But could they serve five times as many customers? No, they could not. There were diminishing returns to labour.

KI 17 p 95

The trouble is that certain factors of production in the shop are fixed:

- The shop is a fixed size. It gets very crowded on a Saturday morning. Assistants sometimes have to wait while customers squeeze past each other to get to the counter, and with five serving, the assistants themselves used to get in each other's way.
- There is only one cash till. Assistants frequently had to wait while other assistants used it.
- There is only one pile of tissue paper for wrapping the bread. Again, the assistants often had to wait.

The fifth and maybe even the fourth assistant ended up serving very few extra customers.

I am still going to the same bread shop and they still have only one till and one pile of tissue paper. But now only three assistants are employed on a Saturday! The shop, however, is just as busy.

How would you advise the baker as to whether he should (a) employ four assistants on a Saturday; (b) extend his shop, thereby allowing more customers to be served on a Saturday morning?

These costs are called **explicit costs** because they involve direct payment of money by firms.

Factors already owned by the firm: implicit costs

When the firm already owns factors (e.g. machinery), it does not as a rule have to pay out money to use them. Their opportunity costs are thus **implicit costs**. They are equal to what the factors could earn for the firm in some alternative use, either within the firm or hired out to some other firm.

Pause for thought

Assume that a farmer decides to grow wheat on land that could be used for growing barley. Barley sells for £100 per tonne. Wheat sells for £150 per tonne. Seed, fertiliser, labour and other costs of growing crops are £80 per tonne for both wheat and barley. What are the farmer's costs and profit per tonne of growing wheat?

Here are some examples of implicit costs:

- A firm owns some buildings. The opportunity cost of using them is the rent it could have received by letting them out to another firm.
- A firm draws £250 000 from the bank out of its savings to invest in new plant and equipment. The opportunity cost of this investment is not just the £250 000 (an explicit cost), but also the interest it thereby forgoes (an implicit cost).
- The owner of the firm could have earned £50 000 per annum by working for someone else. This £50 000, then, is the opportunity cost of the owner's time.

TC 2 p 9

If there is no alternative use for a factor of production, as in the case of a machine designed to make a specific product, and if it has no scrap value, the opportunity cost of using it is *zero*. In such a case, if the output from the machine is worth more than the cost of all the *other* inputs involved, the firm might as well use the machine rather than let it stand idle.

What the firm paid for the machine – its **historic cost** – is irrelevant. Not using the machine will not bring that money back. It has been spent. These are sometimes referred to as **sunk costs in production**.

Costs and output

KI 16 p 94

A firm's costs of production depend on its output. The reason is simple. The more it produces, the greater the quantity of factors of production it must use. The more factors it uses, the greater its costs will be. More precisely, this relationship depends on two elements.

- *The productivity of the factors.* The greater their productivity, the smaller will be the quantity of them that is needed to produce a given level of output, and hence the lower will be the cost of that output.
- *The price of the factors.* The higher their price, the higher will be the costs of production.

In the short run, some factors used by the firm are fixed in supply. Their total costs, therefore, are fixed, in the sense that they do not vary with output. Rent on land is a **fixed cost**. It is the same whether the firm produces a lot or a little.

The total cost of variable factors, however, does vary with output. The cost of raw materials is a **variable cost**. The more that is produced, the more raw materials are used and therefore the higher is their total cost.

Total cost (TC) is thus total fixed cost (TFC) plus total variable cost (TVC).

Average and marginal cost

In addition to total costs (fixed and variable), there are two other categories of costs that are particularly important for our analysis of profits. These are average cost and marginal cost.

Average cost (AC) is cost per unit of production.

$$AC = TC/Q$$

Thus if it costs a firm £20 000 to produce 100 units of a product, the average cost would be £200 for each unit (£20 000/100).

As with total cost, average cost can be divided into two components, fixed and variable. In other words, average cost equals **average fixed cost** ($AFC = TFC/Q$) plus **average variable cost** ($AVC = TVC/Q$).

$$AC = AFC + AVC$$

Definitions

Explicit costs The payments to outside suppliers of inputs.

Implicit costs Costs that do not involve a direct payment of money to a third party, but which nevertheless involve a sacrifice of some alternative.

Historic costs The original amount the firm paid for factors it now owns.

Sunk costs (in production) Costs that cannot be recouped (e.g. by transferring assets to other uses). Examples include specialised machinery or the costs of an advertising campaign.

Fixed costs Total costs that do not vary with the amount of output produced.

Variable costs Total costs that do vary with the amount of output produced.

Total cost The sum of total fixed costs and total variable costs: $TC = TFC + TVC$.

Average (total) cost Total cost (fixed plus variable) per unit of output: $AC = TC/Q = AFC + AVC$

Average fixed cost Total fixed cost per unit of output: $AFC = TFC/Q$

Average variable cost Total variable cost per unit of output: $AVC = TVC/Q$

BOX 5.2 **MALTHUS AND THE DISMAL SCIENCE OF ECONOMICS** CASE STUDIES & APPLICATIONS

Population growth + diminishing returns = starvation

KI 17 p 95

The law of diminishing returns has potentially cataclysmic implications for the future populations of the world.

If the population of the world grows rapidly, then food output may not keep pace with it. There will be diminishing returns to labour as more and more people crowd on to the limited amount of land available.

KI 1 p 5

This is already a problem in some of the poorest countries of the world, especially in sub-Saharan Africa. The land is barely able to support current population levels. Only one or two bad harvests are needed to cause mass starvation – witness the appalling famines in recent years in Ethiopia and the Sudan.

The relationship between population and food output was analysed as long ago as 1798 by the Reverend Thomas Robert Malthus (1766–1834) in his *Essay on the Principle of Population*. This book was a bestseller and made Robert Malthus perhaps the best known of all social scientists of his day.

Malthus argued as follows:

> I say that the power of population is indefinitely greater than the power in the earth to produce subsistence for man.
>
> Population when unchecked, increases in a geometrical ratio. Subsistence increases only in an arithmetical ratio. A slight acquaintance with numbers will show the immensity of the first power in comparison with the second.[1]

What Malthus was saying is that world population tends to double about every 25 years or so if unchecked. It grows geometrically, like the series: 1, 2, 4, 8, 16, 32, 64, etc. But food output, because of diminishing returns, cannot keep pace with this. It is likely to grow at only an arithmetical rate, like the series: 1, 2, 3, 4, 5, 6, 7, etc. It is clear that population, if unchecked, will soon outstrip food supply.

So what is the check on population growth? According to Malthus, it is starvation. As population grows, so food output per head will fall until, with more and more people starving, the death rate will rise. Only then will population growth stabilise at the rate of growth of food output.

Have Malthus's gloomy predictions been borne out by events? Two factors have mitigated the forces that Malthus described:

- The rate of population growth tends to slow down as countries become more developed. Although improved health prolongs life, this tends to be more than offset by a decline in the birth rate as people choose to have smaller families. This is illustrated in the table below. Population growth peaked in the 1960s, has slowed substantially since then and is projected to slow further in future decades.
- Technological improvements in farming have greatly increased food output per hectare. These include better fertilisers and the development of genetically modified crops (see Case Study 5.3 in MyLab Economics for an example.)

The growth in food output has thus exceeded the rate of population growth in developed countries and in some developing countries too. Nevertheless, the Malthusian spectre is very real for some of the poorest developing countries, which are simply unable to feed their populations satisfactorily. It is these poorest countries of the world which have some of the highest rates of population growth – around 3 per cent per annum in many African countries.

A further cause for concern arises from the move in Asia towards a westernised diet, with meat and dairy products playing a larger part. This further increases pressure on the land, since cattle require considerably more grain to produce meat than would be needed to feed humans a vegetarian diet.

A third factor is cited by some commentators, who remain unconvinced of the strength of Malthus's gloomy prognostication for the world. They believe that he seriously underestimated mankind's capacity to innovate; perhaps human ingenuity is one resource that doesn't suffer from diminishing returns (see Section 10.6).

Why might it be possible for there to be a zero marginal productivity of labour on many family farms in poor countries and yet just enough food for all the members of the family to survive?

Construct two charts using data from **World Population Prospects: The 2017 Revision** *(United Nations, Department of Economic and Social Affairs). The first should show the level of the world population since 1950 with forecasts to 2100, and the second the annual rate of growth over the same period. Write a short summary describing what the charts show.*

[1] T.R. Malthus, *First Essay on Population* (Macmillan, 1926), pp. 13–14.

Marginal cost (MC) is the *extra* cost of producing *one more unit*: that is, the rise in total cost per one unit rise in output. Note that all marginal costs are variable, since, by definition, there can be no extra fixed costs as output rises.

$$MC = \frac{\Delta TC}{\Delta Q}$$

where Δ means 'a rise in'.

For example, assume that a firm is currently producing 1 000 000 boxes of matches a month. It now increases output by 1000 boxes (another batch): $\Delta Q = 1000$. Assume that, as a result, total costs rise by £500: $\Delta TC = £500$. What is the cost of producing one more box of matches? It is:

$$MC = \frac{\Delta TC}{\Delta Q} = \frac{£500}{1000} = 50\text{p}$$

Table 5.1 shows costs for an imaginary firm, firm X, over a given period of time (e.g. a week). The table

Definition

Marginal cost The cost of producing one more unit of output: $MC = \Delta TC/\Delta Q$

shows how average and marginal costs can be derived from total costs. It is assumed that total fixed costs are £12 000 (column 2) and that total variable costs are as shown in column 3.

The figures for *TVC* have been chosen to illustrate the law of diminishing returns. Initially, *before* diminishing returns set in, *TVC* rises less and less rapidly as more variable factors are added. For example, in the case of a factory with a fixed supply of machinery, initially as more workers are taken on the workers can do increasingly specialist tasks and make a fuller use of the capital equipment. Above a certain output (3 units in Table 5.1), diminishing returns set in. Given that extra workers (the extra variable factors) are producing less and less extra output, the extra units of output they do produce will be costing more and more in terms of wage costs. Thus *TVC* rises more and more rapidly. You can see this by examining column 3.

The figures in the remaining columns in Table 5.1 are derived from columns 1 to 3. Look at the figures in each of the columns and check how the figures are derived. Note the figures for marginal cost are plotted between the lines to illustrate that marginal cost represents the increase in costs as output increases from one unit to the next. We can use the figures in Table 5.1 to draw *MC*, *AFC*, *AVC* and *AC* curves.

Marginal cost (MC). The shape of the *MC* curve follows directly from the law of diminishing returns. Initially, in Figure 5.1, as more of the variable factor is used, extra units of output cost less than previous units. *MC* falls.

KI 17 p 95

Beyond a certain level of output, diminishing returns set in. This is shown as point *x*. Thereafter *MC* rises. Additional units of output cost more and more to produce, since they require ever-increasing amounts of the variable factor.

Pause for thought

Before you read on, can you explain why the marginal cost curve will always cut the average cost curve at its lowest point?

Average fixed cost (AFC). This falls continuously as output rises, since total fixed costs are being spread over a greater and greater output.

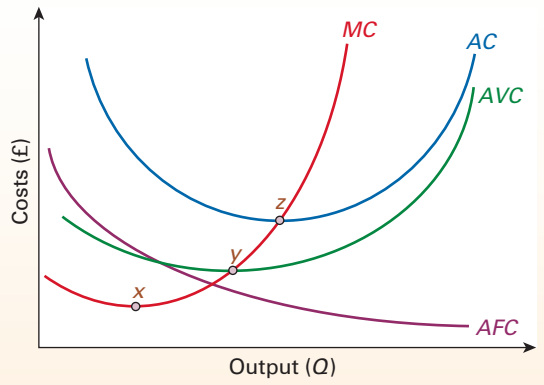
Average (total) cost (AC). The shape of the *AC* curve depends on the shape of the *MC* curve. As long as new units of output cost less than the average, their production must pull the average cost down. That is, if *MC* is less than *AC*, *AC* must be falling. Likewise, if new units cost more than the average, their production must drive the average up. That is, if *MC* is greater than *AC*, *AC* must be rising. Therefore, the *MC* curve crosses the *AC* curve at its minimum point (point *z* in Figure 5.1). This relationship between averages and marginals is explored in Box 5.3.

Average variable cost (AVC). Since $AVC = AC - AFC$, the *AVC* curve is simply the vertical difference between the *AC* and the *AFC* curves. Note that as *AFC* gets less,

Table 5.1 Costs for firm X

Output (*Q*) (1)	*TFC* (£000) (2)	*TVC* (£000) (3)	*TC*(*TFC* + *TVC*) (£000) (4)	*AFC* (*TFC*/*Q*) (£000) (5)	*AVC* (*TVC*/*Q*) (£000) (6)	*AC* (*TC*/*Q*) (£000) (7)	*MC*($\Delta TC/\Delta Q$) (£000) (8)
0	12	0	12	–	–	–	
							10
1	12	10	22	12	10	22	
							6
2	12	16	28	6	8	14	
							5
3	12	21	33	4	7	11	
							7
4	12	28	40	3	7	10	
							12
5	12	40	52	2.4	8	10.4	
							20
6	12	60	72	2	10	12	
							31
7	12	91	103	1.7	13	14.7	

Figure 5.1 Average and marginal costs



Pause for thought

Why is the minimum point of the AVC curve at a lower level of output than the minimum point of the AC curve?

the gap between *AVC* and *AC* narrows. Since all marginal costs are variable (by definition, there are no marginal fixed costs), the same relationship holds between *MC* and *AVC* as it did between *MC* and *AC*. That is, if *MC* is less than *AVC*, *AVC* must be falling, and if *MC* is greater than *AVC*, *AVC* must be rising. Therefore, as with the *AC* curve, the *MC* curve crosses the *AVC* curve at its minimum point (point *y* in Figure 5.1).

Recap

1. In the short run, it is assumed that one or more factors (inputs) are fixed in supply. The actual length of the short run will vary from industry to industry.
2. Production in the short run is subject to diminishing returns. As greater quantities of the variable factor(s) are used, so each additional unit of the variable factor will add less to output than previous units: i.e. output will rise less and less rapidly.
3. When measuring costs of production, we should be careful to use the concept of opportunity cost. In the case of inputs not owned by the firm, the opportunity cost is simply the explicit cost of purchasing or hiring them. It is the price paid for them. In the case of inputs already owned by the firm, it is the implicit cost of what the factor could have earned for the firm in its best alternative use.
4. As some factors are fixed in supply in the short run, their total costs are fixed with respect to output. In the case of variable factors, their total cost increases as more output is produced and hence as more of them are used. Total cost can be divided into total fixed and total variable cost.
5. Marginal cost is the cost of producing one more unit of output. It will probably fall at first but will start to rise as soon as diminishing returns set in.
6. Average cost, like total cost, can be divided into fixed and variable costs. Average fixed cost will decline as more output is produced. The reason is that the total fixed cost is being spread over a greater and greater number of units of output. Average variable cost will tend to decline at first, but once the marginal cost has risen above it, it must then rise. The same applies to average cost.

BOX 5.3 THE RELATIONSHIP BETWEEN AVERAGES AND MARGINALS

EXPLORING ECONOMICS

In this chapter we have just examined the concepts of *average* and *marginal* cost. We shall be coming across several other average and marginal concepts later on. It is useful at this stage to examine the general relationship between averages and marginals. In all cases there are three simple rules that relate them.

To illustrate these rules, consider the following example.

Imagine a room with ten people in it. Assume that the *average* age of those present is 20.

Now if a 20-year-old enters the room (the *marginal* age), this will not affect the average age. It will remain at 20. If a 56-year-old now comes in, the average age will rise: not to 56, of course, but to 23. This is found by dividing the sum of everyone's ages (276) by the number of people (12). If then a child of 10 were to enter the room, this would pull the average age down – to 22(=286/13).

From this example we can derive the three universal rules about averages and marginals:

- If the marginal equals the average, the average will not change.
- If the marginal is above the average, the average will rise.
- If the marginal is below the average, the average will fall.

Suppose a course you are studying has five equally weighted pieces of coursework. Assume each one is marked out of 100 and your results are shown in the table below.

Coursework:	*1*	*2*	*3*	*4*	*5*
Mark:	*60*	*60*	*70*	*70*	*20*

Each number in the second row of the table is the marginal mark from each piece of coursework. Calculate your total and average number of marks after each piece of coursework. Show how the average and marginal marks illustrate the three rules above.

BOX 5.4 **COSTS AND THE ECONOMIC VULNERABILITY OF FIRMS**

The behaviour of costs and firms' financial well-being

The economic environment in which firms operate is highly uncertain and yet, while some firms fail, others grow and succeed. What is it that makes some firms more vulnerable to the economic environment, while other firms are much more insulated? In this box, we look at one aspect of the answer to this question by focusing on both the shape and movements of a firms' average cost curves. How do these affect firms' economic vulnerability?

Type 1 economic vulnerability – average costs and changing levels of output

A typical firm will have a U-shaped average cost curve. It falls at first, reflecting rapidly falling average fixed costs, as they are spread over a greater output, plus a more efficient deployment of variable factors of production. Then, as diminishing marginal returns become relatively more important than falling average fixed costs, average costs rise. An important determinant of how vulnerable a firm is how quickly average costs first fall and then rise as output changes.

Consider two firms, A and B. Each firm's average cost curve is shown in Figure (a). Assume, for simplicity, that each firm achieves minimum average cost at point *x,* namely at the same output Q_0 and at the same average cost, AC_0.

Now consider the impact of COVID-19 and the economic downturn it created. Assume that both firms experienced a fall in demand and that they both reduced output to Q_1.

With a U-shaped *AC* curve, both firms would have experienced a rise in per-unit costs, but firm A's unit costs would have risen significantly faster than firm B's, because firm A has a much steeper *AC* curve. The same fall in quantity would have pushed firm A's average costs up from AC_0 to AC_1 (point *a*) but only caused firm B's costs to increase to AC_2 (point *b*) as firm B's *AC* curve is very flat.

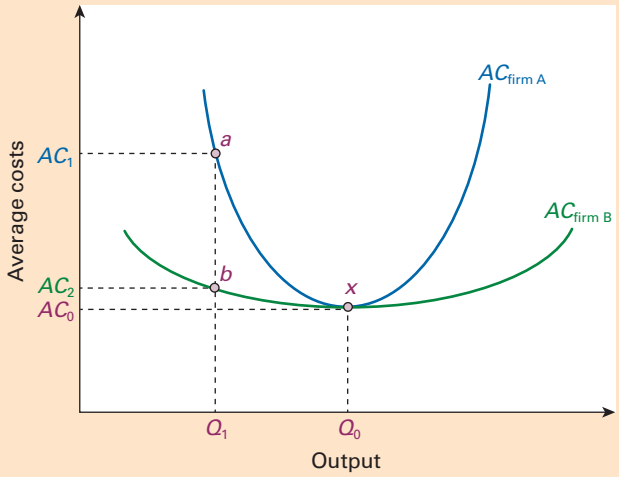
Returning to point *x*, now assume that there is an expansion in demand. Both firms consequently increase output. Again, firm A's costs rise more rapidly than firm B's, because of the shape of the *AC* curves.

Firm A is thus much more susceptible to any change in demand than firm B. Any small decrease (or increase) in output will have a significant effect on firm A's costs and hence on its profit margin and profits, making it a much more vulnerable firm.

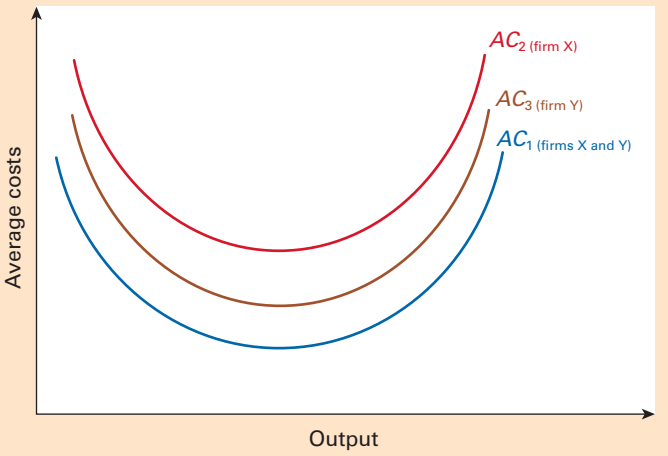
What are the factors that affect the steepness of the *AC* curve and hence make one firm more vulnerable than another?

- *The ratio of fixed factors to variable factors.* The higher the ratio the more likely a firm is to face a steep *AC* curve. Total fixed costs do not change with output and hence, if output falls, it implies that its high fixed costs are being spread over fewer and fewer units of output and this causes average costs to rise rapidly. Many airlines faced this problem during the pandemic.
- *Flexibility in the use of inputs.* A firm that is relatively inflexible in its use of inputs may find that a cut in production results in its average cost rising rapidly. Similarly, if it wishes to expand output beyond Q_0, it may find it difficult to do so without incurring considerable extra costs, for example by employing expensive agency staff or hiring expensive machinery.

(a) *Average cost for firms A and B: change in output*



(b) *Average cost for firms C and D: change in costs*



Type 2 economic vulnerability – average costs and the purchasing of inputs

Most firms purchase inputs to the production process from other companies, but their reliance on other firms and, in some cases on materials where there is a volatile global market, can vary significantly. The second type of economic vulnerability concerns a firm's reliance on external or bought-in inputs.

Some firms can be be heavily dependent on oil or other raw materials and, as such, if the prices of these change, it can have a very big effect on the firm's costs of production and its profit. During periods of rising national and global output, the demand for commodities, such as oil, increases and their prices rise (see Box 2.3). Similarly if there is a fall in their supply, as happened with oil, gas and wheat during the Ukraine war, their prices can rise. A firm that is heavily dependent on commodities can therefore experience a significant effect on its production costs, with its *AC* curve shifting vertically upwards by a considerable amount. Such a firm is thus vulnerable to these input price increases.

By contrast, for a firm whose output is not reliant on commodities during production, changes in global commodity prices may cause only a very small shift in the *AC* curve. Similarly, a firm which can use alternative inputs, when faced with an increase in the price of input A, may simply shift to input B, with little change in its costs. Such a firm is not vulnerable to a rise in price of these inputs.

In Figure (b), assume both firms X and Y have the same *AC* curve, given by AC_1. But let us also assume that firm X is very dependent on oil, whereas firm Y is not.

Assume that oil prices now rise. This leads to a large upward shift in firm X's *AC* curve from AC_1 to AC_2, but a smaller upward shift in firm Y's *AC* curve from AC_1 to AC_3. There is a much larger cost penalty imposed on firm X than on firm Y, due to the reliance on oil as a factor of production.

From late 2014 to 2016 and in 2020, oil prices fell considerably and so those firms which were heavily dependent on oil saw their *AC* curves shift downwards significantly, thereby helping to increase their profits. In 2021–22, oil prices rose, first with economic recovery, and then in 2022 with the war in Ukraine. Consequently, firms that were big users of oil, either directly into the production process or for transporting their inputs and produce, saw their costs rise and their profits eroded. Those firms which were less reliant on oil, however, did not benefit or suffer so much from the changing global price of oil.

The turbulent experience of airlines

Some firms are vulnerable in both ways, being heavily dependent on externally determined costs, while also having a high proportion of fixed costs. This might mean that, as economies move between economic boom and economic recession, these firms are always vulnerable to changes in costs of production.

Good examples are airlines. They have very high fixed costs as a percentage of internal costs, suggesting a steep *AC* curve at low levels of output and hence a vulnerability to a fall in output. However, airlines also require high quantities of fuel and hence are heavily reliant on oil. Thus externally determined costs also account for a high percentage of total costs.

During the COVID-19 pandemic, airlines saw demand fall. EasyJet operated at just 25 per cent of normal capacity; British Airways cancelled 94 per cent of flights; and North American airlines' traffic for 2020 fell 75.4 per cent relative to 2019. Such low demand made airlines very vulnerable to rising short-run average costs. In the following years, airlines saw demand rise, but this too created problems for them, as oil prices rose – a major input for airlines.

Is it possible for a firm to suffer from both type 1 and type 2 economic vulnerability?

Conduct some research on a firm of your choice, looking at information relating to its costs, and discuss whether this firm could suffer from type 1, type 2 or both types of economic vulnerability.

5.2 PRODUCTION AND COSTS: LONG RUN

How do a firm's costs vary with output over the longer term?

KI 16 p 94

In the long run, *all* factors of production are variable. There is time for the firm to build a new factory, to install new machines, to use different production techniques and to combine its inputs in whatever proportion and quantities it chooses.

Looking to the long run, then, a firm will have to make a number of decisions: about the scale of its operations and the techniques of production it will use. These decisions will affect the costs of production. It is important, therefore, to get them right.

The scale of production

If a firm were to double all of its inputs – something it could do in the long run – would it double its output? Or would output more than double or less than double? We can distinguish three possible situations:

Constant returns to scale. This is where a given percentage increase in inputs will lead to the *same* percentage increase in output.

Increasing returns to scale. This is where a given percentage increase in inputs will lead to a *larger* percentage increase in output.

Decreasing returns to scale. This is where a given percentage increase in inputs will lead to a *smaller* percentage increase in output.

Notice the terminology here. The words 'to scale' mean that *all* inputs increase by the same proportion. Decreasing returns to *scale* are therefore quite different from diminishing *marginal* returns (where only the *variable* factor increases). The differences between marginal returns to a variable factor and returns to scale are illustrated in Table 5.2.

In the short run, input 1 is assumed to be fixed in supply (at 3 units). Output can be increased only by using more of the variable factor (input 2). In the long run, however, both input 1 and input 2 are variable.

In the short-run situation, diminishing returns can be seen from the fact that output increases at a decreasing rate (25 to 45 to 60 to 70 to 75) as input 2 is increased. In the long-run situation, the table illustrates increasing returns to scale. Output increases at an *increasing* rate (15 to 35 to 60 to 90 to 125) as both inputs are increased.

KI 17 p 95

Table 5.2 Short-run and long-run increases in output

Short run			Long run		
Input 1	Input 2	Output	Input 1	Input 2	Output
3	1	25	1	1	15
3	2	45	2	2	35
3	3	60	3	3	60
3	4	70	4	4	90
3	5	75	5	5	125

Economies of scale

The concept of increasing returns to scale is closely linked to that of **economies of scale**. A firm experiences economies of scale if costs per unit of output fall as the scale of production increases. Clearly, if a firm is getting increasing returns to scale from its factors of production, then as it produces more it will be using smaller and smaller amounts of factors per unit of output. Other things being equal, this means that it will be producing at a lower average cost.

KI 16 p 94

There are several reasons why firms are likely to experience economies of scale. Some are due to increasing returns to scale; some are not.

Specialisation and division of labour. In large-scale plants, workers can do more simple, repetitive jobs. With this **specialisation and division of labour** less training is needed; workers can become highly efficient in their particular job, especially with long production runs; less time is lost from workers switching from one operation to another; and supervision is easier. Workers and managers can be employed who have specific skills in specific areas.

Indivisibilities. Some inputs are of a minimum size: they are indivisible. The most obvious example is machinery. Take the case of a combine harvester. A small-scale farmer could not make full use of one. They only become economical to use, therefore, on farms above a certain size. The problem of **indivisibilities** is made worse when different machines, each of which is part

Definitions

Economies of scale Where increasing the scale of production leads to a lower cost per unit of output.

Specialisation and division of labour Where production is broken down into a number of simpler, more specialised tasks, thus allowing workers to acquire a high degree of efficiency.

Indivisibilities The impossibility of dividing a factor into smaller units.

of the production process, are of a different size. For example, if there are two types of machine, one producing 6 units a day, the other packaging 4 units a day, a minimum of 12 units per day will have to be produced, involving two production machines and three packaging machines, if all machines are to be fully utilised.

The 'container principle'. Any capital equipment that contains things (e.g. blast furnaces, oil tankers, pipes, vats) tends to cost less per unit of output the larger its size. The reason has to do with the relationship between a container's volume and its surface area. A container's cost depends largely on the materials used to build it and hence roughly on its *surface area*. Its output depends largely on its *volume*. Large containers have a bigger volume relative to surface area than do small containers.

For example, a container with a bottom, top and four sides, with each side measuring 1 metre, has a volume of 1 cubic metre and a surface area of 6 square metres (6 surfaces of 1 square metre each). If each side were doubled in length to 2 metres, the volume would be 8 cubic metres and the surface area 24 square metres (6 surfaces of 4 square metres each). Thus an eightfold increase in capacity has been gained at only a fourfold increase in the container's surface area, and hence an approximate fourfold increase in cost.

Greater efficiency of large machines. Large machines may be more efficient in the sense that more output can be gained for a given amount of inputs. For example, only one worker may be required to operate a machine, whether it be large or small. Also, a large machine may make more efficient use of raw materials.

By-products. With production on a large scale, there may be sufficient waste products to make some by-product.

Multi-stage production. A large factory may be able to take a product through several stages in its manufacture. This saves time and cost moving the semi-finished product from one firm or factory to another. For example, a large cardboard-manufacturing firm may be able to convert trees or waste paper into cardboard and then into cardboard boxes in a continuous sequence.

All the above are examples of **plant economies of scale**. They are due to an individual factory or workplace or machine being large. There are other economies of scale that are associated with the *firm* being large – perhaps with many factories.

Organisational economies. With a large firm, individual plants can specialise in particular functions. There can also be centralised administration of the firms. Often, after a merger between two firms, savings can be made by **rationalising** their activities in this way.

Spreading overheads. There are some expenditures that are only economic when the *firm* is large, such as research and development: only a large firm can afford to set up a research laboratory. This is another example of indivisibilities, only this time at the level of the firm rather than the plant. The greater the firm's output, the more these **overhead costs** are spread.

Financial economies. Large firms can often obtain finance at lower interest rates than small firms, as they are perceived as having lower default risks or have more power to negotiate a better deal. Additionally, they may be able to obtain certain inputs more cheaply by buying in bulk. (These are examples of economies of scale which are not the result of increasing returns to scale.)

Economies of scope. Often a firm is large because it produces a range of products. This can result in each individual product being produced more cheaply than if it was produced in a single-product firm.

The reason for these **economies of scope** is that various overhead costs and financial and organisational economies can be shared among the products. For example, a firm such as Apple that produces mobile phones, tablets, computers, smart watches, etc. can benefit from shared marketing and distribution costs and the bulk purchase of electronic components.

Diseconomies of scale

When firms get beyond a certain size, costs per unit of output may start to increase. There are several reasons for such **diseconomies of scale**:

Pause for thought

Which of the economies of scale we have considered are due to increasing returns to scale and which are due to other factors?

Definitions

Plant economies of scale Economies of scale that arise because of the large size of the factory.

Rationalisation The reorganising of production (often after a merger) so as to cut out waste and duplication and generally to reduce costs.

Overhead costs Costs arising from the general running of an organisation, and only indirectly related to the level of output.

Economies of scope Where increasing the range of products produced by a firm reduces the cost of producing each one.

Diseconomies of scale Where costs per unit of output increase as the scale of production increases.

- Management problems of co-ordination may increase as the firm becomes larger and more complex, and as lines of communication get longer. There may be a lack of personal involvement by management.
- Workers may feel 'alienated' if their jobs are boring and repetitive, and if they feel an insignificantly small part of a large organisation. Poor motivation may lead to shoddy work.
- Industrial relations may deteriorate as a result of these factors and also as a result of the more complex inter-relationships between different categories of worker.
- Production-line processes and the complex inter-dependencies of mass production can lead to great disruption if there are hold-ups in any one part of the firm.

Pause for thought

Why are firms likely to experience economies of scale up to a certain size and then diseconomies of scale after some point beyond that?

Whether firms experience economies or diseconomies of scale depends on the conditions applying in each individual firm.

The size of the whole industry

As an *industry* grows in size, this can lead to **external economies of scale** for its member firms. This is where a firm, whatever its own individual size, benefits from the *whole industry* being large. For example, the firm may benefit from having access to specialist raw material or component suppliers, labour with specific skills, firms that specialise in marketing the finished product, and banks and other financial institutions with experience of the industry's requirements.

What we are referring to here is the **industry's infrastructure**: the facilities, support services, skills and experience that can be shared by its members. This is one reason why we often see industrial clusters emerge.

The member firms of a particular industry might, however, experience **external diseconomies of scale**. For example, as an industry grows larger, this may create a growing shortage of specific raw materials or skilled labour. This will push up their prices, and hence the firms' costs.

Long-run average cost

We turn now to *long-run* cost curves. Since there are no fixed factors in the long run, there are no long-run fixed costs. For example, the firm may rent more land in order to expand its operations. Its rent bill therefore goes up as it expands its output. All costs, then, in the long run are variable costs.

Although it is possible to draw long-run total, marginal and average cost curves, we will concentrate on **long-run average cost (*LRAC*) curves**. These can take various shapes, but a typical one is shown in Figure 5.2.

It is often assumed that as a firm expands, it will initially experience economies of scale and thus face a downward-sloping *LRAC* curve. After a point, however, all such economies will have been achieved and thus the curve will flatten out. Then, possibly after a period of constant *LRAC*, the firm will get so large that it will start experiencing diseconomies of scale and thus a rising *LRAC*. At this stage, production and financial economies begin to be offset by the managerial problems of running a giant organisation.

The effect of this is to give a saucer-shaped curve, as in Figure 5.2.

Assumptions behind the long-run average cost curve

We make three key assumptions when constructing long-run average cost curves:

Definitions

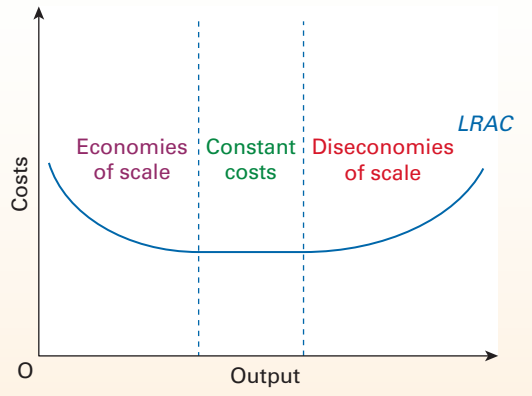
External economies of scale Where a firm's costs per unit of output decrease as the size of the whole *industry* grows.

Industry's infrastructure The network of supply agents, communications, skills, training facilities, distribution channels, specialised financial services, etc. that supports a particular industry.

External diseconomies of scale Where a firm's costs per unit of output increase as the size of the whole industry increases.

Long-run average cost curve A curve that shows how average cost varies with output on the assumption that all factors are variable. (It is assumed that the least-cost method of production will be chosen for each output.)

Figure 5.2 A typical long-run average cost curve



Factor prices are given. At each level of output it is assumed that a firm will be faced with a given set of factor prices. If factor prices *change*, therefore, both short- and long-run cost curves will shift. Thus an increase in nationally negotiated wage rates would shift the curves upwards.

However, factor prices might be different at *different* levels of output. For example, one of the economies of scale that many firms enjoy is the ability to obtain bulk discount on raw materials and other supplies. In such cases the curve does *not* shift. The different factor prices are merely experienced at different points along the curve and are reflected in the shape of the curve. Factor prices are still given for any particular level of output.

The state of technology and factor quality are given. These are assumed to change only in the *very* long run (see page 106–7). If a firm gains economies of scale, it is because it is exploiting *existing* technologies and/or making better use of the existing availability of factors of production. As technology improves the curves will shift downwards.

Firms choose the least-cost combination of factors for each output. The assumption here is that firms operate efficiently: that they choose the cheapest possible way of producing any level of output. Efficiency in the sense of firms combining factors optimally is

BOX 5.5 THE OPTIMUM COMBINATIONS OF INPUTS

EXPLORING ECONOMICS

Equi-marginal principle in production

TC 2 p 9

We saw in Box 4.2 how we can apply marginal utility analysis to understand better how a rational consumer decides what combination of goods to buy. We use an equivalent marginal approach to examine how a profit-maximising firm will allocate inputs to production so as employ the least costly combination.

Making production choices at the margin

TC 2 p 9

The equi-marginal principle underpins rational decision making. When comparing alternative choices, to maximise net benefit, the ratio of marginal benefit to marginal cost should be the same for all alternatives. So with n alternatives, net benefit is maximised where:

$$\frac{MB_a}{MC_a} = \frac{MB_b}{MC_b} = \frac{MB_c}{MC_c} \cdots = \frac{MB_n}{MC_n} \quad (1)$$

In terms of production, the marginal benefit is known as the ***marginal physical product*** (*MPP*). This is the *extra* output produced by employing *one more* unit of the input. In the case of inputs of labour (L) and capital (K), for example, we can refer to the marginal physical product of labour (MPP_L) and the marginal physical product of capital (MPP_K). The former measures the additional output from using one more unit of labour, while the latter measures the additional output from using one more unit of capital.

When a firm uses just labour and capital, the least-cost combination of the two will be where the *MPP* of each factor relative to its price (P) is equal. This can be written as:

$$\frac{MPP_L}{P_L} = \frac{MPP_K}{P_K} \quad (2)$$

If they were not equal, it would be possible to reduce cost per unit of output by using a different combination of labour and capital. For example, if:

$$\frac{MPP_L}{P_L} > \frac{MPP_K}{P_K} \quad (3)$$

then more labour should be used relative to capital, since the firm is getting a greater physical return for its money from extra workers than from extra capital. As more labour is used per unit of capital, however, diminishing returns to labour take effect. Thus, MPP_L will fall. Likewise, as less capital is used per unit of labour, MPP_K will rise. This will continue until equation (2) is satisfied. At this point, the firm will stop substituting labour for capital. This new combination of factors or 'choice of technique' will now be the most productively efficient.

KI 17 p 95

The same principle applies when a firm uses many different factors. If we assume a firm uses n factors, the least-cost combination of factors will be where:

$$\frac{MPP_a}{P_a} = \frac{MPP_b}{P_b} = \frac{MPP_c}{P_c} \cdots = \frac{MPP_n}{P_n} \quad (4)$$

If any inequality exists between the MPP/P ratios, a firm will be able to reduce its costs by using more of those factors with a high MPP/P ratio and less of those with a low MPP/P ratio until they all become equal.

A major problem for a firm in choosing the least-cost technique is in predicting future factor price changes. If the price of a factor were to change, the MPP/P ratios would no longer be equal. To minimise costs the firm would then like to alter its factor combinations until the MPP/P ratios once more became equal.

The trouble is that, once it has committed itself to a particular technique, it may be several years before it can switch to an alternative one. Thus, if a firm invests in labour-intensive methods of production and is then faced with an unexpected wage rise, it may regret not having chosen a more capital-intensive technique.

While there is no simple solution to this issue, there are a number of companies that have made a business of predicting technological and price trends across different sectors to assist firms in their decision making.

? *If factor X costs twice as much as factor Y ($P_X/P_Y = 2$), what can be said about the relationship between the MPPs of the two factors if the optimum combination of factors is used?*

Definition

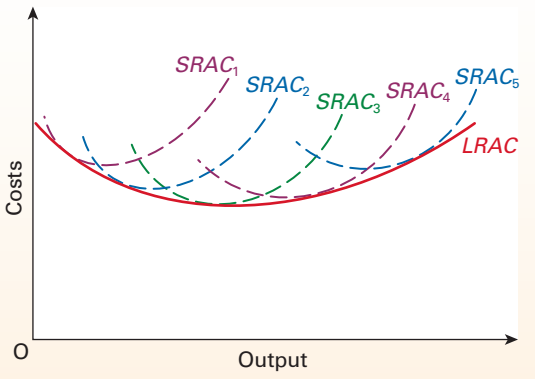
Marginal physical product The additional output from using one more unit of an input.

known as **productive efficiency**. Firms that are allocating factors optimally are unable to alter factor combinations without raising costs.

The optimal allocation of factors can be understood by applying the *equi-marginal principle*. This is the principle we applied to rational consumers choosing the combination of goods and services to maximise their utility (see pages 75–6). A firm will choose a combination of inputs where the marginal benefit per pound (£) spent is equal across all inputs. If this were not the case, it could reduce the average cost of its output by using a different combination of inputs. This is explored further in Box 5.5 and Web Appendix 5.2.

The key point is that, if a firm does not operate efficiently, it would be producing at a point *above* the *LRAC* curve.

Figure 5.3 Constructing long-run average cost curves from short-run average cost curves



The relationship between long-run and short-run average cost curves

Take the case of a firm that has just one factory and faces a short-run average cost curve, illustrated by $SRAC_1$ in Figure 5.3.

In the long run, it can build more factories. If it thereby experiences economies of scale (due, say, to savings on administration), each successive factory will allow it to produce with a new lower *SRAC* curve. Thus with two factories it will face curve $SRAC_2$; with three factories curve $SRAC_3$, and so on. Each *SRAC* curve corresponds to a particular amount of the factor that is fixed in the short run: in this case, the factory. (Many more *SRAC* curves could be drawn between the ones shown, since factories of different sizes could be built or existing ones could be expanded.)

From this succession of short-run average cost curves we can construct a long-run average cost curve. This is shown in Figure 5.3. This is known as the **envelope curve**, since it envelops the short-run curves.

Long-run cost curves in practice

Firms do experience economies of scale. Some experience continuously falling *LRAC* curves; others experience economies of scale up to a certain output and thereafter constant returns to scale.

Evidence is inconclusive on the question of diseconomies of scale. There is little evidence to suggest the existence of *technical* diseconomies, but the possibility of diseconomies due to managerial and industrial relations problems cannot be ruled out.

Some evidence on economies of scale in the UK and the EU is considered in Box 5.6.

Postscript: decision- making in different time periods

We have distinguished between the short run and the long run. Let us introduce two more time periods to complete the picture. The complete list then reads as follows.

Very short run (immediate run). *All* factors are fixed. Output is fixed. The supply curve is vertical. On a day-to-day basis a firm may not be able to vary output at all. For example, a flower seller, once the day's flowers have been purchased from the wholesaler, cannot alter the number of flowers available for sale on that day. In the very short run, all that may remain for a producer to do is to sell an already produced good.

Short run. At least one factor is fixed in supply. More can be produced, but the firm will come up against the law of diminishing returns as it tries to do so. KI 17 p 95

Long run. All factors are variable. The firm may experience constant, increasing or decreasing returns to scale. But although all factors can be increased or decreased, they are of a fixed *quality*.

Very long run. All factors are variable, *and* their quality and hence productivity can change. Labour productivity can increase as a result of education,

Definitions

Productive efficiency Producing with the least-cost combination of factors for a given output.

Envelope curve A long-run average cost curve drawn as the tangency points of a series of short-run average cost curves.

training, experience and social factors. The productivity of capital can increase as a result of new inventions (new discoveries) and innovation (putting inventions into practice).

Improvements in factor quality will reduce costs and thus shift the short- and long-run cost curves downwards.

Just how long the 'very long run' is will vary from firm to firm. It will depend on how long it takes to develop new techniques, new skills or new work practices.

It is important to realise that decisions for all four time periods can be made at the same time. Firms do not make short-run decisions in the short run and long-run decisions in the long run. They can make both short-run and long-run decisions today. For example, assume that a firm experiences an increase in consumer demand and anticipates that it will continue into the foreseeable future. It thus wants to increase output. Consequently, it makes the following four decisions today:

- *Very short run*. It accepts that for a few days it will not be able to increase output. It informs its customers that they will have to wait. It may temporarily raise prices to choke off some of the demand.
- *Short run*. It negotiates with labour to introduce overtime working as soon as possible, to tide it over the next few weeks. It orders extra raw materials from its suppliers. It launches a recruitment drive for new labour so as to avoid paying overtime longer than is necessary.
- *Long run*. It starts proceedings to build a new factory. The first step may be to discuss requirements with a firm of consultants.
- *Very long run*. It institutes a programme of research and development and/or training in an attempt to increase productivity.

Although we distinguish these four time periods, it is the middle two we are primarily concerned with. The reason for this is that there is very little the firm can do in the *very* short run. And in the *very* long run, although the firm will obviously want to increase the productivity of its inputs, it will not be in the position to make precise calculations of how to do it. It will not know precisely what inventions will be made, or just what will be the results of its own research and development.

Pause for thought

Why are Christmas trees and fresh foods often sold cheaply on Christmas Eve? Why do shops find it most profitable to lower the price of these items to the point where the price elasticity of demand equals −1?

Recap

1. In the long run, a firm is able to vary the quantity it uses of all factors of production. There are no fixed factors and hence no long-run fixed costs.
2. If it increases all factors by the same proportion, it may experience constant, increasing or decreasing returns to scale.
3. Economies of scale occur when costs per unit of output fall as the scale of production increases. This can be the result of a number of factors, some of which are directly due to increasing (physical) returns to scale. These include the benefits of specialisation and division of labour, the use of larger and more efficient machines, and the ability to have a more integrated system of production. Other economies of scale arise from the financial and administrative benefits of large-scale organisations.
4. When constructing long-run cost curves it is assumed that factor prices are given, that the state of technology is given and that firms will choose the least-cost method of production for each given output.
5. The *LRAC* curve can be downward sloping, upward sloping or horizontal, depending in turn on whether there are economies of scale, diseconomies of scale or neither. Typically *LRAC* curves are drawn as saucer-shaped (or as l-shaped). As output expands, initially there are economies of scale. When these are exhausted the curve will become flat. When the firm becomes very large it may begin to experience diseconomies of scale. If this happens, the *LRAC* curve will begin to slope upwards again.
6. An envelope curve can be drawn which shows the relationship between short-run and long-run average cost curves. The *LRAC* curve envelops the short-run *AC* curves: it is 'tangential' to them (i.e. just touches them).
7. To minimise costs per unit of output, a firm should choose that combination of factors which gives an equal marginal product for each factor relative to its price. If the *MPP*/*P* ratio for one factor is greater than for another, then more of this factor should be used until the *MPP* on the last pound (£) spent across all factors is equal.
8. Four distinct time periods can be distinguished. In addition to the short- and long-run periods, we can also distinguish the very short- and very long-run periods. The very short run is when all factors are fixed. The very long run is where not only the quantity of factors but also their quality is variable (as a result of changing technology, etc.).

BOX 5.6 MINIMUM EFFICIENT SCALE

The extent of economies of scale in practice

Two of the most important studies of economies of scale have been those made by C.F. Pratten[1] in the late 1980s and by a group advising the European Commission[2] in 1997. Both studies found strong evidence that many firms, especially in manufacturing, experienced substantial economies of scale.

In a few cases long-run average costs fell continuously as output increased. For most firms, however, they fell up to a certain level of output and then remained constant.

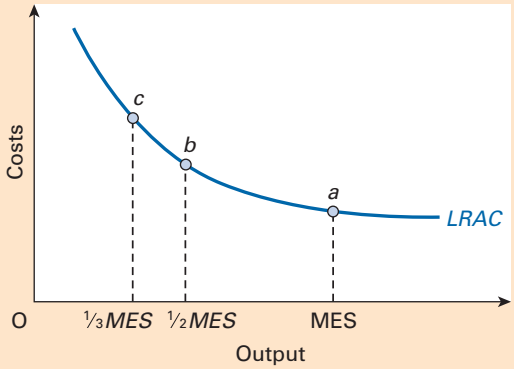
The extent of economies of scale can be measured by looking at a firm's *minimum efficient scale (MES)*. The *MES* is the size beyond which no significant additional economies of scale can be achieved: in other words, the point where the *LRAC* curve flattens off. In Pratten's studies he defined this level as the minimum scale above which any possible doubling in scale would reduce average costs by less than 5 per cent (i.e. virtually the bottom of the *LRAC* curve). In the diagram *MES* is shown at point *a*.

The *MES* can be expressed in terms either of an individual factory or of the whole firm. Where it refers to the minimum efficient scale of an individual factory, the *MES* is known as the *minimum efficient plant size* (*MEPS*).

The *MES* can then be expressed as a percentage of the total size of the market or of total domestic production. Table (a), based on the Pratten study, shows *MES* for plants and firms in various industries. The first column shows *MES* as a percentage of total UK production. The second column shows *MES* as a percentage of total EU production. Table (b), based on the 1997 study, shows *MES* for various plants as a percentage of total EU production.

Expressing *MES* as a percentage of total output gives an indication of how competitive the industry could be. In some industries (such as footwear and carpets), economies of scale were exhausted (i.e. *MES* was reached) with plants or firms that were still small relative to total UK production and

Long-run average cost curve



even smaller relative to total EU production. In such industries there would be room for many firms and thus scope for considerable competition.

In other industries, however, even if a single plant or firm were large enough to produce the whole output of the industry in the UK, it would still not be large enough to experience the full potential economies of scale: the *MES* is greater than 100 per cent. Examples from Table (a) include factories producing cellulose fibres, and car manufacturers. In such industries there is no possibility of competition. In fact, as long as the *MES* exceeds 50 per cent there will not be room for more than one firm large enough to gain full economies of scale. In this case the industry is said to be a *natural monopoly*. As we shall see in the next few chapters, when competition is lacking, consumers may suffer by firms charging prices considerably above costs.

A second way of measuring the extent of economies of scale is to see how much costs would increase if production were reduced to a certain fraction of *MES*. The normal fractions used are $\frac{1}{2}$ or $\frac{1}{3}$ *MES*. This is illustrated in the diagram. Point *b* corresponds to $\frac{1}{2}$ *MES*; point *c* to $\frac{1}{3}$ *MES*. The greater the percentage by which *LRAC* at point *b* or *c* is higher than at point *a*, the greater will be the economies of scale to be

[1]C.F. Pratten, 'A survey of the economies of scale', in *Research on the 'Costs of Non-Europe'*, Vol. 2 (Office for Official Publications of the European Communities, Luxembourg, 1988).

[2]European Commission/Economists Advisory Group Ltd, 'Economies of Scale', *The Single Market Review, Sub-series V, Vol. 4*. (Office for Official Publications of the European Communities, Luxembourg, 1997).

5.3 REVENUE

How does a firm's revenue vary with its level of sales?

Remember that we defined a firm's total profit as its total revenue minus its total costs of production. In the last two sections we looked at costs. We now turn to revenue.

As with costs, we distinguish between three revenue concepts: total revenue (*TR*), average revenue (*AR*) and marginal revenue (*MR*).

Total, average and marginal revenue

Total revenue (TR)

Total revenue is the firm's total earnings per period of time from the sale of a particular amount of output (Q).

For example, if a firm sells 1000 units (Q) per month at a price of £5 each (P), then its monthly

CASE STUDIES & APPLICATIONS

Table (a)

Product	*MES* as % of production		% additional cost at $^1/_2$ *MES*
	UK	EU	
Individual plants			
Cellulose fibres	125	16	3
Rolled aluminium semi-manufactures	114	15	15
Refrigerators	85	11	4
Steel	72	10	6
Electric motors	60	6	15
TV sets	40	9	9
Cigarettes	24	6	1.4
Ball-bearings	20	2	6
Beer	12	3	7
Nylon	4	1	12
Bricks	1	0.2	25
Tufted carpets	0.3	0.04	10
Shoes	0.3	0.03	1
Firms			
Cars	200	20	9
Lorries	104	21	7.5
Mainframe computers	>100	[n.a.]	5
Aircraft	100	[n.a.]	5
Tractors	98	19	6

Source: See footnote 1 above

gained by producing at *MES* rather than at $^1/_2$ *MES* or $^1/_3$ *MES*. For example, in Table (a) greater economies of scale can be gained by moving from $^1/_2$ *MES* to *MES* in the production of electric motors than of cigarettes.

Table (b)

Plants	*MES* as % of total EU production
Aerospace	12.19
Tractors and agricultural machinery	6.57
Electric lighting	3.76
Steel tubes	2.42
Shipbuilding	1.63
Rubber	1.06
Radio and TV	0.69
Footwear	0.08
Carpets	0.03

Source: See footnote 2 above

The main purpose of the studies was to determine whether the single EU market is big enough to allow both economies of scale and competition. The tables suggest that in all cases, other things being equal, the EU market is indeed large enough for this to occur. The second study also found that 47 of the 53 manufacturing sectors analysed had scope for further exploitation of economies of scale.

1. *Why might a firm operating with one plant achieve MEPS and yet not be large enough to achieve MES? (Clue: are all economies of scale achieved at plant level?)*
2. *Why might a firm producing bricks have an MES that is only 0.2 per cent of total EU production and yet face little effective competition from other EU countries?*

Conduct a short literature search looking into the relationship between international trade and economies of scale. Summarise your findings and the theories behind this relationship.

total revenue will be £5000: in other words, £5 × 1000($P \times Q$). Thus:

$$TR = P \times Q$$

Average revenue (AR)

Average revenue is the amount that the firm earns per unit sold. Thus:

$$AR = TR/Q$$

So if the firm earns £5000 (TR) from selling 1000 units (Q), it will earn £5 per unit. But this is simply the price! Thus:

$$AR = P$$

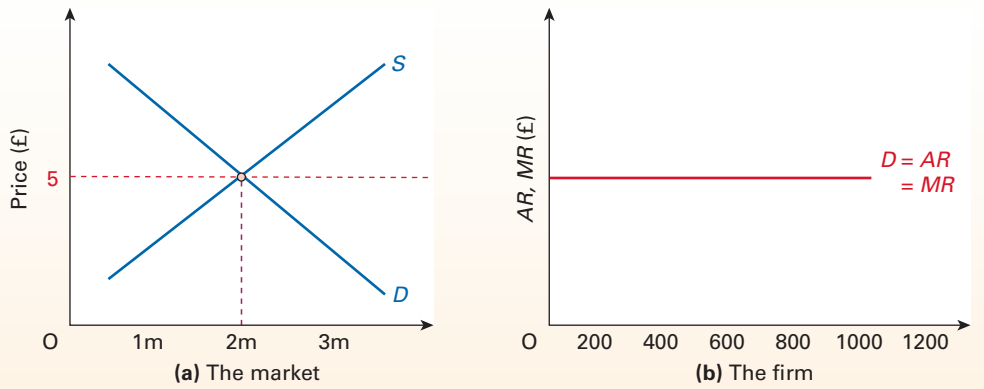
(The only exception to this is when the firm is selling its products at different prices to different consumers.

Definitions

Total revenue A firm's total earnings from a specified level of sales within a specified period: $TR = P \times Q$

Average revenue Total revenue per unit of output. When all output is sold at the same price, average revenue will be the same as price: $AR = TR/Q = P$

Figure 5.4 Deriving a firm's *AR* and *MR*: price-taking firm



In this case *AR* is simply the (weighted) average price.)

Marginal revenue (MR)

Marginal revenue is the extra total revenue gained by selling one more unit (per time period). So if a firm sells an extra 20 units this month compared with what it expected to sell, and in the process earns an extra £100, then it is getting an extra £5 for each extra unit sold: $MR = £5$. Thus:

$$MR = \Delta TR/\Delta Q$$

We now need to see how each of these three revenue concepts (*TR*, *AR* and *MR*) varies with output. We can show this relationship graphically in the same way as we did with costs.

The relationship will depend on the market conditions under which a firm operates. A firm that is too small to be able to affect market price will have different-looking revenue curves from a firm that is able to choose the price it charges. Let us examine each of these two situations in turn.

Average and marginal revenue curves when price is not affected by the firm's output

Average revenue

If a firm is very small relative to the whole market, it is likely to be a **price taker**. That is, it has to accept the price given by the intersection of demand and supply in the whole market. But, being so small, it can sell as much as it is capable of producing at that price. This is illustrated in Figure 5.4.

Diagram (a) shows market demand and supply. Equilibrium price is £5. Diagram (b) looks at the demand for an individual firm that is tiny relative to the whole market. (Look at the difference in the scale of the horizontal axes in the two diagrams.)

Being so small, any change in its output will be too insignificant to affect the market price. It thus faces a horizontal demand 'curve' at this price. It can sell 200 units, 600 units, 1200 units or whatever without affecting this £5 price.

Average revenue is thus constant at £5. The firm's average revenue curve must therefore lie along exactly the same line as its demand curve.

Marginal revenue

In the case of a horizontal demand curve, the marginal revenue curve will be the same as the average revenue curve, since selling one more unit at a constant price (*AR*) merely adds that amount to total revenue. If an extra unit is sold at a constant price of £5, an extra £5 is earned.

Average and marginal revenue curves when price varies with output

Rather than accepting (or taking) the market price, firms would generally prefer to be a **price maker**. This will be the case when a firm faces a downward-sloping demand curve. Generally, the larger its share of the market, the less elastic will be its demand curve, as there are probably fewer substitutes.

Definitions

Marginal revenue The extra revenue gained by selling one more unit per period of time: $MR = \Delta TR/\Delta Q$

Price taker A firm that is too small to be able to influence the market price.

Price maker A firm that can choose the price it charges; it faces a downward-sloping demand curve. If, however, it alters its price, this will affect the quantity sold: a fall in price will lead to more being sold; a higher price will lead to less.

As a price maker, the firm can choose to lower its price and thereby sell more. Alternatively, it could raise its price, and earn more per unit sold, if it was willing to accept a fall in sales. Firms will tend to be benefit from being price makers, as is discussed in an article from Harvard Business School.[1]

The curves (*AR* and *MR*) will look quite different when price does vary with the firm's output.

Average revenue

Remember that average revenue equals price. If, therefore, the price has to be reduced to sell more output, average revenue will fall as output increases.

Table 5.3 gives an example of a firm facing a downward-sloping demand curve. The demand curve (which shows how much is sold at each price) is given by the first two columns.

Note that, as in the case of a price-taking firm, the demand curve and the *AR* curve lie along exactly the same line (see Figure 5.5). The reason for this is simple: $AR = P$, and thus the curve relating price to quantity (the demand curve) must be the same as that relating average revenue to quantity (the *AR* curve).

Marginal revenue

When a firm faces a downward-sloping demand curve, marginal revenue will be less than average revenue, and may even be negative. But why?

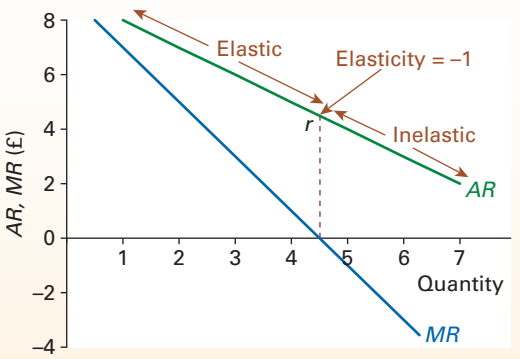
If a firm is to sell more per time period, it must lower its price (assuming it does not advertise). This will mean lowering the price not just for the extra units it hopes to sell, but also for those units it would have sold had it not lowered the price.

Thus the marginal revenue is the price at which it sells the last unit, *minus* the loss in revenue it has incurred by reducing the price on those units it could otherwise have sold at the higher price. This can be illustrated with Table 5.3.

Table 5.3 Revenues for a firm facing a downward-sloping demand curve

Q (units)	*P* = *AR* (£)	*TR* (£)	*MR* (£)
1	8	8	
			6
2	7	14	
			4
3	6	18	
			2
4	5	20	
			0
5	4	20	
			−2
6	3	18	
			−4
7	2	14	

Figure 5.5 *AR* and *MR* curves for a firm facing a downward-sloping demand curve



Assume that price is currently £7. Two units are thus sold. The firm now wishes to sell an extra unit. It lowers the price to £6. But it must lower the price on all units sold, as it cannot tell who will pay £7 and who will only pay £6. It thus gains £6 from the sale of the third unit, but loses £2 by having to reduce the price by £1 on the two units it could otherwise have sold at £7. Its net gain is therefore £6 − £2 = £4. This is the marginal revenue: it is the extra revenue gained by the firm from selling one more unit. Try using this method to check out the remaining figures for *MR* in Table 5.3. (Note that in the table the figures for *MR* are entered in the spaces between the figures for the other three columns.)

There is a simple relationship between marginal revenue and *price elasticity of demand*. Remember from Section 3.2 (pages 52–5) that if demand is price elastic, a *decrease* in price will lead to a proportionately larger increase in the quantity demanded and hence to an *increase* in revenue. Marginal revenue will thus be positive. If, however, demand is inelastic, a decrease in price will lead to a proportionately smaller increase in sales. In this case the price reduction will more than offset the increase in sales and as a result revenue will fall. Marginal revenue will be negative.

TC 10 p 49

If, then, marginal revenue is a positive figure (i.e. if sales per time period are 4 units or less in Figure 5.5), the demand curve will be elastic at that quantity, since a rise in quantity sold (as a result of a reduction in price) would lead to a rise in total revenue. If, on the other hand, marginal revenue is negative (i.e. at a level of sales of 5 or more units in Figure 5.5), the demand curve will be inelastic at that quantity, since a rise in quantity sold would lead to a *fall* in total revenue.

Thus the demand (*AR*) curve of Figure 5.5 is elastic to the left of point *r* and inelastic to the right.

[1] B.P. Shapiro, 'Commodity busters: be a price maker not a price taker', *Working Knowledge* (Harvard University, 10/2/03).

Shifts in revenue curves

We saw in Chapter 2 that a change in *price* will cause a movement along a demand curve. It is similar with revenue curves, except that here the causal connection is in the other direction. Here we ask what happens to revenue when there is a change in the firm's *output*. Again the effect is shown by a movement along the curves.

A change in any *other* determinant of demand, such as tastes, income or the price of other goods, will shift the demand curve. As this change affects the price at which each level of output can be sold, there will be a shift in all three revenue curves. An increase in revenue is shown by a vertical shift upwards; a decrease by a shift downwards.

Recap

1. Total revenue (*TR*) is the total amount a firm earns from its sales in a given time period. It is simply price times quantity: $TR = P \times Q$.
2. Average revenue (*AR*) is total revenue per unit: $AR = TR/Q$. In other words, $AR = P$.
3. Marginal revenue (*MR*) is the extra revenue earned from the sale of one more unit per time period: $MR = \Delta TR/\Delta Q$.
4. The *AR* curve will be the same as the demand curve for the firm's product. In the case of a price taker, the demand curve and hence the *AR* curve will be a horizontal straight line and will also be the same as the *MR* curve.
5. A firm that faces a downward-sloping demand curve must obviously also face the same downward-sloping *AR* curve. The *MR* curve will also slope downwards, but will be below the *AR* curve and steeper than it.
6. When demand is price elastic, marginal revenue will be positive. When demand is price inelastic, marginal revenue will be negative.
7. A change in output is represented by a movement along the revenue curves. A change in any other determinant of revenue will shift the curves up or down.

5.4 PROFIT MAXIMISATION

How much output should a firm produce if it wants to maximise its profit?

We are now in a position to put costs and revenue together to find the output at which profit is maximised, and also to find out how much that profit will be. We take the case of a firm facing a downward-sloping demand curve, whose revenue, costs and profit are shown in Table 5.4.

Finding the maximum profit that the firm can make is a two-stage process. The first stage is to find the profit-maximising output. To do this we use the *MC* and *MR* curves. The second stage is to find out just how much profit is at this output. To do this we use the *AR* and *AC* curves.

Stage 1: Using marginal curves to arrive at the profit-maximising output

TC 2 p 9

There is a very simple **profit-maximising rule**: if profits are to be maximised, *MR must equal MC*. From Table 5.4 it can be seen that $MR = MC$ at an output of 3. This is shown as point e in Figure 5.6.

But why are profits maximised when $MR = MC$? The simplest way of answering this is to see what the position would be if *MR* did not equal *MC*.

Referring to Figure 5.6, at a level of output below 3, *MR* exceeds *MC*. This means that by producing more units there will be a bigger addition to revenue (*MR*) than to cost (*MC*). Total profit will *increase*. *As long as MR exceeds MC, profit can be increased by increasing production.*

At a level of output above 3, *MC* exceeds *MR*. All levels of output above 3 thus add more to cost than to revenue and hence *reduce* profit. *As long as MC exceeds MR, profit can be increased by cutting back on production.*

Profits are thus maximised where $MC = MR$: at an output of 3. This can be confirmed by reference to the *TΠ* column in Table 5.4.

Students worry sometimes about the argument that profits are maximised when $MR = MC$. Surely, they say, if the last unit is making no profit, how can profit be at a *maximum*? The answer is very simple. If you cannot add anything more to a total, the total must be

Definition

Profit-maximising rule Profit is maximised where marginal revenue equals marginal cost.

Table 5.4 Revenue, cost and profit

Q (units)	*P* = *AR* (£)	*TR* (£)	*MR* (£)	*TC* (£)	*AC* (£)	*MC* (£)	*TΠ* (£)	*AΠ* (£)
0	9	0		6	–		−6	–
			8			4		
1	8	8		10	10		−2	−2
			6			2		
2	7	14		12	6		2	1
			4			2		
3	6	18		14	4.67		4	1.33
			2			4		
4	5	20		18	4.5		2	0.5
			0			7		
5	4	20		25	5		−5	−1
			−2			11		
6	3	18		36	6		−18	−3
			−4			20		
7	2	14		56	8		−42	−6

at the maximum. Take the simple analogy of going up a hill. When you cannot go any higher, you must be at the top.

Stage 2: Using average curves to measure the size of the profit

Once the profit-maximising output has been discovered, we now use the average curves to measure the *amount* of profit at the maximum. Both marginal and average curves corresponding to the data in Table 5.4 are plotted in Figure 5.7

First, average profit ($A\Pi$) is found. This is simply $AR - AC$. At the profit-maximising output of 3, this gives a figure for ($A\Pi$) of £6 − £$4^2/_3$ = £$1^1/_3$. Then total profit is obtained by multiplying average profit by output:

$$T\Pi = A\Pi \times Q$$

This is shown as the shaded area. It equals £$1^1/_3 \times 3$ = £4. This can again be confirmed by reference to the $T\Pi$ column in Table 5.4.

Figure 5.6 Finding the profit-maximising output using marginal curves

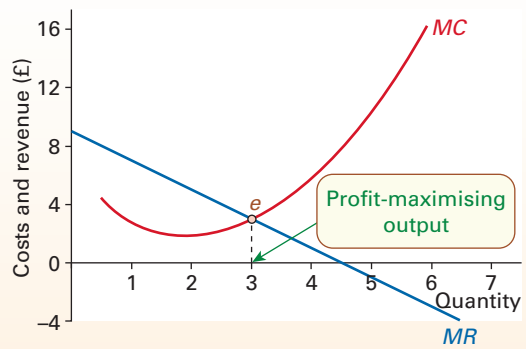
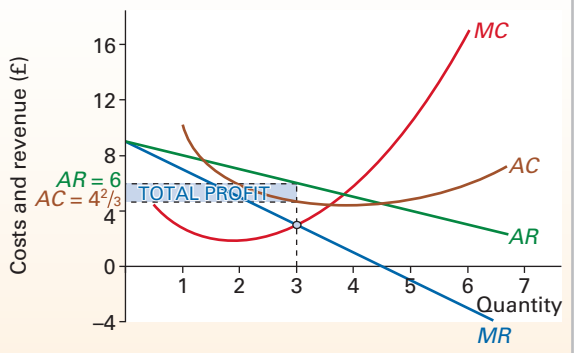


Figure 5.7 Measuring the maximum profit using average curves



Pause for thought

What will be the effect on a firm's profit-maximising output of a rise in fixed costs?

Some qualifications

Long-run profit maximisation

Assuming that the *AR* and *MR* curves are the same in the long run as in the short run, long-run profits will be maximised at the output where *MR* equals the *long-run MC*. The reasoning is the same as with the short-run case.

The meaning of 'profit'

One element of cost is the opportunity cost to the owners of the firm incurred by being in business. This is the minimum return that the owners must make on their capital in order to prevent them from eventually deciding to close down and perhaps move into some alternative business. It is a *cost* since, just as with wages, rent, etc., it has to be covered if the firm is to continue producing. This opportunity cost to the owners is sometimes known as **normal profit**, and is *included in the cost curves*.

TC 1 p 7

What determines this normal rate of profit? It has two components. First, someone setting up in business invests capital in it. There is thus an opportunity cost. This is the interest that could have been earned by lending it in some riskless form (e.g. by putting it in a savings account in a bank). Nobody would set up a business unless they

Definition

Normal profit The opportunity cost of being in business. It consists of the interest that could be earned on a riskless asset, plus a return for risk-taking. It is counted as a cost of production.

Figure 5.8 Loss-minimising output

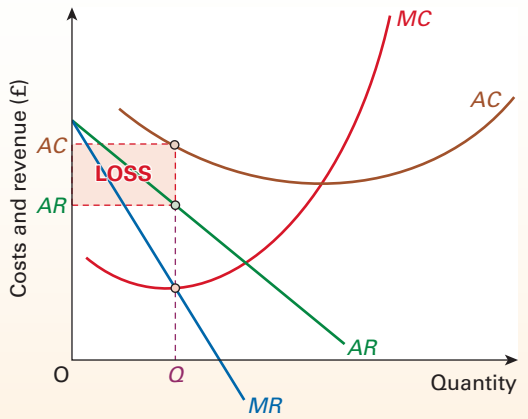
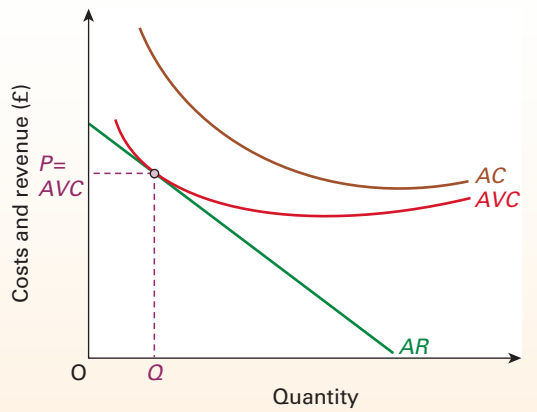


Figure 5.9 The short-run shut-down point



expected to earn at least this rate of profit. Running a business is far from riskless, however, and hence a second element is a return to compensate for risk. Thus:

Normal profit(%) = Rate of interest on a risk-less loan + A risk premium

The risk premium varies according to the line of business. In those businesses with fairly predictable patterns, such as food retailing, it is relatively low. Where outcomes are very uncertain, as in mineral exploration or the manufacture of fashion garments, it is relatively high.

Thus if owners of a business earn normal profit, they will (just) be content to remain in that industry. If they earn more than normal profit, they will also (obviously) prefer to stay in this business. If they earn less than normal profit, then after a time they will consider leaving and using their capital for some other purpose.

Given that normal profits are included in costs, any profit that is shown diagrammatically (e.g. the shaded area in Figure 5.7) must therefore be over and above normal profit. It is known by several alternative names: **supernormal profit**, pure profit, economic profit, abnormal profit, producer's surplus (or sometimes simply profit). They all mean the same thing: the excess of total profit over normal profit, or *TR* minus *TC*.

Loss minimising

Sometimes there is no output at which the firm can make a profit. Such a situation is illustrated in Figure 5.8: the *AC* curve is above the *AR* curve at all levels of output.

In this case, the output where $MR = MC$ will be the loss-minimising output. The amount of loss at the point where $MR = MC$ is shown by the shaded area in Figure 5.8.

Whether or not to produce at all

The short run. Fixed costs have to be paid even if the firm is producing nothing at all. Rent has to be paid, business rates have to be paid, and so forth. Providing, therefore, that the firm is more than covering its *variable* costs, it can go some way to paying off these fixed costs and therefore will continue to produce.

It will shut down if it cannot cover its variable costs: that is, if the *AVC* curve is above, or the *AR* curve is below, that illustrated in Figure 5.9. This situation is known as the **short-run shut-down point**.

The long run. All costs are variable in the long run. If, therefore, the firm cannot cover its long-run average costs (which include normal profit), it will close down. The **long-run shut-down point** will be where the *AR* curve is tangential to (i.e. just touches) the *LRAC* curve.

Definitions

Supernormal profit (also known as pure profit, economic profit, abnormal profit, producer's surplus or simply profit). The excess of total profit above normal profit.

Short-run shut-down point Where the *AR* curve is tangential to the *AVC* curve. The firm can only just cover its variable costs. Any fall in revenue below this level will cause a profit-maximising firm to shut down immediately.

Long-run shut-down point Where the *AR* curve is tangential to the *LRAC* curve. The firm can just make normal profits. Any fall in revenue below this level will cause a profit-maximising firm to shut down once all costs have become variable.

Recap

1. Total profit equals total revenue minus total cost.
2. Graphically, profits are maximised at the output where marginal revenue equals marginal cost. Having found this output, the level of maximum profit can be found by finding the average profit ($AR - AC$) and then multiplying it by the level of output.
3. Normal profit is the minimum profit that must be made to persuade a firm to stay in business in the long run. It is counted as part of the firm's costs. Supernormal profit is any profit over and above normal profit.
4. For a firm that cannot make a profit at any level of output, the point where $MR = MC$ represents the loss-minimising output.
5. In the short run, a firm will close down if it cannot cover its variable costs. In the long run, it will close down if it cannot make normal profits.

5.5 PROBLEMS WITH TRADITIONAL THEORY

Behavioural economics of the firm

The traditional approach to modelling firms' behaviour that we have considered to this point assumes that they aim to maximise profits. Although this is an accurate assumption for many firms, for many it is not.

Some firms would *like* to maximise profits, but have insufficient information to enable them to do so. Others do not even want to maximise profits, if that means sacrificing achieving some other aim, such as rapid growth or increased market share.

Many firms, especially larger ones, are complex organisations, with different individuals and departments pursuing their own agenda. What happens when these various goals come into conflict? How does conflict get resolved? What are the implications for consumers and other 'stakeholders'?

Explaining actual producer behaviour

Lack of information

The main difficulty in trying to maximise profits is a *lack of information*. Firms operate in very complex environments, dealing with large amounts of imperfect information and uncertainty about both the present and the future.

One consequence is that firms are unlikely to know precisely the nature of their demand and *MR* curves. While firms will know how much they are selling at the moment, this only gives them one point on their demand curve and no point at all on their *MR* curve. In order to make even an informed guess of marginal revenue, they must have some idea of how responsive demand will be to a change in price. But how are they to estimate this price elasticity? While market research may help, it is frequently unreliable.

TC 10 p 49

But there is also the problem that firms operate in a changing environment where cost and revenue curves shift. If a firm chooses a price and output that maximises profits this year, it may as a result jeopardise profits in the future. Therefore, firms often face the problem of deciding the time period over which they should be seeking to maximise profits.

For example, a firm may be considering whether to invest in new expensive equipment which results in costs rising in the short run and thus short-run profits fall. However, if the quality of the product increases, demand is likely to increase over the longer run. Furthermore, its variable costs are likely to decrease if the new equipment is more efficient. Consequently, long-run profit is likely to increase, but probably by a highly uncertain amount.

Heuristics

Trying to work out and implement a profit-maximising strategy in a complex environment, dealing with imperfect information and uncertainty about both the present and the future, is a cognitively demanding task. Hence, rather like the consumers we discussed in Section 4.4, managers may respond by using *heuristics*: rules of thumb or mental short-cuts to simplify things. Some heuristics might include the following:

Simple pricing rules. In the case of **average cost or mark-up pricing**, producers work out the price by

Definition

Average cost or mark-up pricing Where firms set the price by adding a profit mark-up to average cost.

simply adding a certain percentage (mark-up) for profit on top of average costs (average fixed costs plus average variable costs).

Copying the strategy of the most profitable businesses in the market. This is only possible if a firm can observe the actions and profits made by its rivals. But sometimes, this is relatively simple. For example, a rival may engage in price cutting or aggressive advertising – strategies that are relatively easy to copy. Under certain circumstances, though, this could lead to more intense competition, with lower prices and higher output generally.

Focusing on relative rather than absolute profits. It is easier to see if a firm is making more profit than its competitors than if it is making the maximum profit possible. Therefore, managers may judge their success by comparing their firm's performance *relative* to that of their rivals. There may also be financial incentives to behave in this way as bonuses are often based on relative rather than absolute profits.

This type of behaviour can lead to firms implementing strategies that reduce their own profits (e.g. an aggressive price war) if it reduces the profits of their rivals by a greater amount.

Making a satisfactory or target level of profit. Instead of constantly looking for new opportunities to maximise profits in a dynamic market, managers may instead only change a firm's strategy when its profits fall below some target level or if shareholders perceive its profits to be too low.

Over-optimism

Economists have long identified over-optimism as a trait seen in many people. Adam Smith commented that 'the chance of gain is by every man, more or less over-valued, and the chance of loss is by most men under-valued'.[2] Consequently, some firms are likely to be adventurous and prepared to take risks. Adventurous firms are most likely to be those dominated by a powerful and ambitious individual – an individual who is risk loving and therefore prepared to take gambles.

The more dispersed the decision-making power is in a firm, however, and the more worried managers are about their own survival, the more cautious are their policies likely to be. They may, for example, prefer to stick with what have been popular products, and to expand steadily. If a firm is too cautious, however, it may not survive. It may find that it loses markets to more aggressive competitors.

Pause for thought

Why might it present problems for a firm if managers are overconfident? Can you think of any reasons why managers might be more inclined to optimism than most other people?

The principal–agent problem

The traditional theory of the firm assumes that it is the *owners* of the firm who make price and output decisions. It is reasonable to assume that owners *will* want to maximise profits. But, consider the case of **public limited companies** (see Case Study 5.1 on the student website). The owners are the shareholders. While they will probably want the firm to maximise profits to increase their dividends and the value of their shares, they are not directly in control of the firm. They elect directors who, in turn, employ professional managers. It is these professional managers who are decision makers; it is they who are in control. There is thus a *separation between the ownership and control* of a firm.

This therefore raises questions about the objectives of managers. Will *they* want to maximise profits, or will they have some other aim, for example higher salaries, greater power or prestige, better working conditions, greater sales, etc.?

The example of the public limited company illustrates a general feature of any organisation: people (principals) have to employ others (agents) to carry out their wishes. Firms frequently employ consultants to give them advice, or engage the services of specialist firms such as an advertising agency.

Even employees of a company can be seen as 'agents' of their employer. In the case of workers, they can be seen as the agents of management. Junior managers are the agents of senior management. Senior managers are the agents of the directors, who are themselves agents of the shareholders. Thus in large firms there is often a complex chain of principal–agent relationships.

Principal–agent relationships create a potential problem since agents may not act in the best interests of principals. Hence, this problem is known as the

Definition

Public limited company A company owned by its shareholders. Shareholders' liability is limited to the value of their shares. Shares may be bought and sold publicly – on the stock market.

[2]See: Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* Book 1, Chapter 10, para. 29. Adam Smith's *An Inquiry into the Nature and Causes of the Wealth of Nations* was first published in 1776. This edition of Smith's work is based on Edwin Cannan's careful 1904 compilation (Methuen and Co., Ltd) of Smith's fifth edition of the book (1789), the final edition in Smith's lifetime.

principal–agent problem. At the heart of this problem is an asymmetry of information between the principal and the agent. The agent tends to know more than the principal: that is often why they are employed. But it can mean that agents may often get away with operating in their own interests to the detriment of the principal.

KEY IDEA 18

Principal–agent problem Where people (principals), as a result of a lack of knowledge, cannot ensure that their best interests are served by their agents. Agents may take advantage of this asymmetry of information to the disadvantage of the principals.

This explains why firms have mechanisms for *monitoring* the performance of workers, such as annual performance reviews. Furthermore, there may be *incentives* for agents to behave in the principals' interests. For example, managers' salaries or bonuses could be closely linked to the firm's profitability.

TC 6 p 24

Pause for thought

Is the problem of a separation of ownership and control relevant in the case of a state-owned enterprise?

Alternative aims

Profit satisficing

Despite informational asymmetries, agents still need to satisfy their principals. Managers, for example, may need to ensure that *sufficient* profits are made to keep owners happy, but that may be very different from *maximising* profits. Decision makers in many large firms are thus **profit satisficers**. That is, managers strive hard for a minimum target level of profit, but are less interested in profits above this level.

In large firms with many departments (sales, production, design, purchasing, personnel, finance, etc.) it is likely that each will have its own specific set of aims and objectives. This results in a firm having multiple goals. These goals may come into conflict with each other and in turn be constrained by the interests of shareholders, workers, customers and creditors (collectively known as **stakeholders**), who will need to be kept sufficiently happy. It is argued that firms with multiple goals are more likely to be *satisficers* with targets perhaps changed fairly infrequently.

Sales revenue maximisation

Perhaps the most well-known alternative theory of the firm is that of **sales revenue maximisation**. It may be that the success of managers, and in particular sales managers, is judged according to the level of the firm's sales. Sales figures are an obvious barometer of the firm's health. Managers' salaries, power and prestige may depend directly on sales revenue. The firm's sales representatives may be paid commission on their sales. Thus sales revenue maximisation may be a more dominant aim in the firm than profit maximisation, particularly if it has a dominant sales department.

In the case of a sales revenue maximiser, rather than obeying the '$MR = MC$ rule' of profit maximisation, the firm will continue to expand output until producing more units ceases to increase revenue. In other words, it will produce output until the marginal revenue from the last unit is zero, i.e. $MR = 0$ (assuming that at least normal profits are made at this output).

Pause for thought

What is the value of the price elasticity of demand at the sales revenue maximising output?

An important consequence of producing up to the point that $MR = 0$ is that a sales revenue maximiser produces a *greater* level of output than a profit maximiser. This is because it will continue to increase output so long as this increases total revenue even when additional units increase costs by more than they increase revenue. You can see this by inspecting Figure 5.6 on page 113. Profit is maximised at an output of 3; sales revenue is maximised at an output between 4 and 5.

Pause for thought

Draw a diagram with MC *and* MR *curves. Mark the output (a) at which profits are maximised; (b) at which sales revenue is maximised.*

Definitions

Principal–agent problem Where people (principals), as a result of lack of knowledge, cannot ensure that their best interests are served by their agents.

Profit satisficing Where decision makers in a firm aim for a target level of profit rather than the absolute maximum level.

Stakeholders (in a company) People who are affected by a company's activities and/or performance (customers, employees, owners, creditors, people living in the neighbourhood, etc.). They may or may not be in a position to take decisions, or influence decision taking, in the firm.

Sales revenue maximisation An alternative theory which assumes that managers aim to maximise the firm's short-run total revenue.

Long-run profit maximisation

Another alternative theory of the firm worthy of particular consideration is that of **long-run profit maximisation**. The traditional theory of the firm is based on the assumption of *short-run* profit maximisation. Many actions of firms may be seen to conflict with this aim and yet could be consistent with the aim of long-run profit maximisation. For example, policies to increase the size of the firm or the firm's share of the market may involve heavy advertising or low prices to the detriment of short-run profits. But if this results in a larger market share, the resulting economic power may enable the firm to make larger profits in the long run.

At first sight, a theory of long-run profit maximisation would seem to be a realistic alternative to the traditional short-run profit-maximisation theory. In practice, however, the theory is not a very useful predictor of firms' behaviour and is very difficult to test. A claim by managers that they were attempting to maximise long-run profits could be consistent with virtually any policy.

Even if long-run profit maximisation *is* the prime aim, the means of achieving it are extremely complex. Current pricing and marketing decisions affect future demand, while current investment decisions affect future costs. The resulting shifts in demand and cost curves will be very difficult to estimate with any precision. Quite apart from this, the actions of competitors, suppliers, unions, the future state of the economy, and so on, are difficult to predict. Thus it is unrealistic to think of firms making precise calculations of long-run profit-maximising prices and outputs.

It may be appropriate in some circumstances, however, to view firms, when making current price, output and investment decisions, as accounting for the approximate effect on new entrants, consumer demand, future costs, etc. Therefore, they may try to avoid decisions that would appear to conflict with long-run profits.

TC 11 p 61

Definition

Long-run profit maximisation An alternative theory which assumes that managers aim to shift cost and revenue curves so as to maximise profits over some longer time period.

Recap

1. Traditional theory assumes profit-maximising behaviour by firms. Two major criticisms of this are: (a) firms may not have the information to maximise profits; (b) they may not even want to maximise profits.
2. The complexity of the environment in which firms operate may mean that firms adopt simple mental short cuts or 'rules of thumb' known as heuristics. Examples include applying a mark-up over costs when determining prices, imitating other firms or focusing on the firm's relative performance compared with that of competitors.
3. In many companies there is likely to be a separation between ownership and control. It is the managers who make the decisions, and managers may look to maximise their own utility rather than that of the owners (e.g. shareholders). The problem of managers not pursuing the same goals as the owners is an example of the principal–agent problem.
4. Firms are characterised by a series of principal–agent relationships because of asymmetries of information between various parties. Firms may use monitoring mechanisms, such as performance reviews, and incentive mechanisms, such as aligning financial bonuses to performance measures to help align better the interests of various parties.
5. In large firms, decisions are taken by or influenced by a number of different people, including various managers, shareholders, workers, customers, suppliers and creditors. If these different people have different aims, a conflict between them is likely to arise. A firm cannot maximise more than one of any conflicting aims. The alternative is to seek to achieve a satisfactory target level of a number of aims.
6. Managers may gain utility from maximising sales revenue. They will, however, still have to ensure that a satisfactory level of profit is achieved. The output of a firm which seeks to maximise sales revenue will be higher than that for a profit-maximising firm.
7. Rather than seeking to maximise short-run profits, a firm may take a longer-term perspective. It is very difficult, however, to predict the behaviour of a long-run profit-maximising firm, since (a) different managers are likely to make different judgements about how to achieve maximum profits and (b) demand and cost curves may shift unpredictably, both in response to the firm's own policies and as a result of external factors.

QUESTIONS

1. Up to roughly how long is the short run in the following cases?
   a. A firm supplying DJs for clubs and parties.
   b. Nuclear power generation.
   c. A street food wagon.
   d. 'Superstore Hypermarkets plc'.

   In each case specify your assumptions.
2. Given that there is a fixed supply of land in the world, what implications can you draw from the law of diminishing returns about the effects of an increase in world population on food output per head?
3. The following are some costs incurred by a shoe manufacturer. Decide whether each one is a fixed cost or a variable cost or has some element of both.
   a. The cost of leather.
   b. The fee paid to an advertising agency.
   c. Wear and tear on machinery.
   d. Business rates on the factory.
   e. Electricity for heating and lighting.
   f. Electricity for running the machines.
   g. Basic minimum wages agreed with the union.
   h. Overtime pay.
   i. Depreciation of machines as a result purely of their age (irrespective of their condition).
4. What economies of scale is a large department store likely to experience?
5. Why are many firms likely to experience economies of scale up to a certain size and then diseconomies of scale after some point beyond that?
6. Name some industries where external economies of scale are gained. What are the specific external economies in each case?
7. Examine Figure 5.2 (on page 104). What would (i) the firm's long-run total cost curve, and (ii) its long-run marginal cost curve look like?
8. Under what circumstances is a firm likely to experience a flat-bottomed *LRAC* curve?
9. Draw a downward-sloping demand curve. Now choose scales for both axes. Read off various points on the demand curve and use them to construct a table showing price and quantity. Use this table to work out the figures for a marginal revenue column. Now use these figures to draw an *MR* curve.
10. Copy Figure 5.5 (which is based on Table 5.3). Now assume that incomes have risen and that, as a result, two more units per time period can be sold at each price. Construct a new table and plot the resulting new *AR* and *MR* curves on your diagram. Are the new curves parallel to the old ones? Explain.
11. Using the data in Table 5.3, construct a *total revenue* curve and mark the parts of the curve where the price elasticity of demand is (a) elastic; (b) unit elastic; (c) inelastic.
12. From the information given in the following table, construct a table like Table 5.4.

Q	0	1	2	3	4	5	6	7
P	12	11	10	9	8	7	6	5
TC	2	6	9	12	16	21	28	38

13. Use your table to draw a diagram like Figure 5.7. Use this diagram to show the profit-maximising output and the level of maximum profit. Confirm your findings by reference to the table you have constructed.
14. Normal profits are regarded as a cost (and are included in the cost curves). Explain why.
15. What determines the size of normal profit? Will it vary with the general state of the economy?
16. A firm will continue producing in the short run even if it is making a loss, providing it can cover its variable costs. Explain why. Just how long will it be willing to continue making such a loss?
17. The price of pocket calculators and digital watches fell significantly in the years after they were first introduced, and at the same time demand for them increased substantially. Use cost and revenue diagrams to illustrate these events. Explain the reasoning behind the diagram(s) you have drawn.
18. The table below shows the average cost and average revenue (price) for a firm at each level of output.

Output	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
AC (£)	7.00	5.00	4.00	3.30	3.00	3.10	3.50	4.20	5.00	6.00
AR (£)	10.00	9.50	9.00	8.50	8.00	7.50	7.00	6.50	6.00	5.50

   a. Construct a table to show *TC, MC, TR* and *MR* at each level of output (put the figures for *MC* and *MR* midway between the output figures).
   b. Using *MC* and *MR* figures, find the profit-maximising output.
   c. Using *TC* and *TR* figures, check your answer to (b).
   d. Plot the *AC, MC, AR* and *MR* figures on a graph.
   e. Mark the profit-maximising output and the *AR* and *AC* at this output. Shade in an area to represent the level of profits at this output.
19. Would it be possible for firms to calculate their maximum-profit output if they did not use marginal cost and marginal revenue concepts?
20. What is meant by the principal–agent problem? Give two examples of this problem that you have come across in your own experience.
21. 'A firm will always prefer to make more profit rather than less.' Do you agree with this statement? Is it compatible with alternatives to the profit-maximising theory of the firm?
22. Why is it difficult to test the assumption that firms seek to maximise *long-run* profits?
23. Do behavioural theories of the firm allow us to make any predictions about firms' prices and output?
24. Are 'special offers' likely to benefit consumers?

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

5.1 **The legal and organisational structure of firms.** This case looks at the distinction between sole proprietorships, partnerships, private limited companies and public limited companies. It also distinguishes between U-form, M-form and H-form organisations.

5.2 **Division of labour in a pin factory.** This is the famous example of division of labour given by Adam Smith in his *Wealth of Nations* (1776).

5.3 **Diminishing returns to nitrogen fertiliser.** This case study provides a good illustration of diminishing returns in practice by showing the effects on grass yields of the application of increasing amounts of nitrogen fertiliser.

5.4 **The fallacy of using historic costs.** This looks at the example of the pricing of Christmas trees.

5.5 **Followers of fashion.** This case study examines the effects of costs on prices of fashion-sensitive goods.

5.6 **Cost curves in practice.** What do cost curves look like when fixed factors are divisible?

5.7 **Putting on a duplicate.** This examines the effects on marginal costs of additional passengers on a coach journey.

5.8 **The logic of logistics.** This case examines how logistics (the inflow of resources to a company and the outflow of finished goods from it) can be managed to help drive down costs or drive up revenues.

5.9 **How firms increase profits by understanding irrational consumers.** This looks at how firms can increase profits by taking consumer 'irrationality' into account.

WEB APPENDICES

5.1 **The short-run production function.** An analysis of how output varies with increases in the quantity of the variable factor used. The appendix looks at the concepts of total, average and marginal product.

5.2 **The optimum combination of factors.** This appendix builds on the analysis in the text to show how factors can be combined to minimise costs for any given output. It uses the concept of marginal product and develops the 'equi-marginal principle' (see Web Appendix 4.1 for the application of the principle to consumption).

5.3 **Isoquant analysis.** This develops a model to show the optimum combination of factors to minimise costs for any given output or maximise output for any given cost. The analysis is similar to indifference analysis, developed in Web Appendix 4.2.

Chapter 6

Market structures

As we saw in Chapter 5, a firm's profits are maximised where its marginal cost equals its marginal revenue: $MC = MR$. But we will want to know more than this.

Assuming that a firm does want to maximise profits, what determines the *amount* of profit that it will make? Will its profits be large, or just enough for it to survive, or so low that it will be forced out of business? Will the price charged to the consumer be high or low? And, more generally, will the consumer benefit from the decisions a firm makes?

The answers to these questions depend on the amount of competition that a firm faces. A firm in a highly competitive environment will behave quite differently from a firm facing little or no competition. In particular, a firm facing competition from many other firms will be forced to keep its prices down and be as efficient as possible, simply to survive.

Even if a firm faces only one or two rivals, competition might be quite intense. Firms might put a lot of effort into producing more efficiently or into developing new or better products in order to gain a larger share of the market. They may, however, collude with each other to keep prices up.

When firms face little or no competition (like the local water company or a major pharmaceutical company), they may have considerable power over prices and we may end up paying a lot more.

In this chapter we look at different types of market and how well they serve the consumer.

After studying this chapter, you should be able to answer the following questions:

- What determines the degree of market power of a firm?
- Why does operating under conditions of perfect competition make being in business a constant battle for survival?
- How do firms get to become monopolies and remain so?
- At what price and output will a monopolist maximise profits and how much profit will it make?
- How well or badly do monopolies serve the consumer compared with competitive firms?
- How are firms likely to behave when there are just a few of them competing ('oligopolies')? Will they engage in all-out competition or will they collude with each other?
- What strategic games are oligopolists likely to play in their attempt to out-do their rivals? What determines the output of these 'games'?
- Why may firms charge different consumers different prices for an identical product?

6.1 THE DEGREE OF COMPETITION

How much competition does a firm face?

It is traditional to divide industries into categories according to the degree of competition that exists between the firms within the industry. There are four such categories.

At one extreme is **perfect competition** where there are very many firms competing. Each firm is so small relative to the whole industry that it has no power to influence price. It is a price taker. At the other extreme is **monopoly**, where there is just one firm in the industry, and hence no competition from *within* the industry. In the middle comes **monopolistic competition**, which involves quite a lot of firms competing and where there is freedom for new firms to enter the industry, and **oligopoly**, where there are only a few firms and where entry of new firms is restricted.

To distinguish more precisely between these four categories, the following must be considered:

- How freely firms can enter the industry. Is entry free or restricted? If it is restricted, just how great are the barriers to the entry of new firms?
- The nature of the product. Do all firms produce an identical product, or do firms produce their own particular brand or model or variety?
- TC 10 p 49 The degree of control the firm has over price. Is the firm a price taker or can it choose its price, and if so, how will changing its price affect its profits? What we are talking about here is the nature of the demand curve it faces. How elastic is it? If the firm puts up its price, will it lose (a) all its sales (a horizontal demand curve), or (b) a large proportion of its sales (a relatively elastic demand curve), or (c) just a small proportion of its sales (a relatively inelastic demand curve)?

Table 6.1 shows the differences between the four categories.

KEY IDEA 19

Market power benefits the powerful at the expense of others. **When firms have market power over prices, they can use this to raise prices and profits above the perfectly competitive level. Other things being equal, the firm will gain at the expense of the consumer. Similarly, if consumers or workers have market power they can use this to their own benefit.**

Definitions

Perfect competition A market structure where there are many firms; where there is freedom of entry into the industry; where all firms produce an identical product; and where all firms are price takers.

Monopoly A market structure where there is only one firm in the industry.

Monopolistic competition A market structure where, like perfect competition, there are many firms and freedom of entry into the industry, but where each firm produces a differentiated product and thus has some control over its price.

Oligopoly A market structure where there are few enough firms to enable barriers to be erected against the entry of new firms.

Table 6.1 Features of the four market structures

Type of market	Number of firms	Freedom of entry	Nature of product	Examples	Implication for demand curve for firm
Perfect competition	Very many	Unrestricted	Homogeneous (undifferentiated)	Cabbages, carrots, foreign exchange (these *approximate* to perfect competition)	Horizontal. The firm is a price taker
Monopolistic competition	Many/several	Unrestricted	Differentiated	Builders, restaurants, hairdressers, garage mechanics	Downward sloping, but relatively elastic. The firm has some control over price
Oligopoly	Few	Restricted	1. Undifferentiated or 2. Differentiated	1. Petrol 2. Cars, electrical appliances, supermarkets, retail banking	Downward sloping, relatively inelastic but depends on reactions of rivals to a price change
Monopoly	One	Restricted or completely blocked	Unique	Prescription drugs produced under a patent, local water companies	Downward sloping, more inelastic than oligopoly. The firm has considerable control over price

The market structure under which a firm operates will determine its behaviour. Firms under perfect competition behave quite differently from firms that are monopolists, which behave differently again from firms under oligopoly or monopolistic competition.

This behaviour (or 'conduct') in turn affects the firm's performance: its prices, profits, efficiency, etc. In many cases it also affects other firms' performance: *their* prices, profits, efficiency, etc. The collective conduct of all the firms in the industry affects the whole industry's performance.

Economists thus see a causal chain running from market structure to the performance of that industry.

Structure → Conduct → Performance

This does not mean, however, that all firms operating in a particular market structure will behave in exactly the same way. It is for this reason that government policy towards firms – known as 'competition policy' – prefers to focus on the *conduct* of individual firms, rather than simply on the market structure within which they operate.

The conduct of firms can actually affect the development of the market structure. For example, the interaction between firms may influence the development of new products or new production methods, and may encourage or discourage the entrance of new firms into the industry.

It is also important to note that some firms with different divisions and products may operate in more than one market structure. As an example, consider the case of Microsoft. Its Edge browser competes with more successful rivals, such as Chrome and Safari and, as a result, it has little market power in the browser market. Its Office products, by contrast, have a much bigger market share and dominate the word processing, presentation and spreadsheet markets with Word, PowerPoint and Excel, respectively.

In looking at how market structure influences firms' behaviour we begin our analysis with the two extreme market structures: perfect competition and monopoly (Sections 6.2 and 6.3). Then we turn to look at the two intermediate cases of monopolistic competition and oligopoly (Sections 6.4 and 6.5).

These two intermediate cases are sometimes referred to collectively as **imperfect competition**. The vast majority of firms in the real world operate under imperfect competition.

It is still worth studying the two extreme cases, however, because they provide a framework within which to understand the real world. Some industries tend more to the competitive extreme, and thus their performance corresponds to some extent to perfect competition. Other industries tend more to the other extreme: for example, when there is one dominant firm and a few much smaller firms. In such cases their performance corresponds more to monopoly.

Pause for thought

Give one more example in each of the four market categories in Table 6.1.

Recap

1. There are four alternative market structures under which firms operate. In ascending order of firms' market power, they are: perfect competition, monopolistic competition, oligopoly and monopoly.
2. The market structure under which a firm operates affects its conduct, which in turn affects its performance.

6.2 PERFECT COMPETITION

What happens when there are very many firms all competing against each other? Is this good for us as consumers?

The theory of perfect competition illustrates an extreme form of capitalism. In it, firms are entirely subject to market forces. They have no power whatsoever to affect the price of the product. The price they face is determined by the interaction of demand and supply in the whole *market*.

Assumptions

The model of perfect competition is built on four assumptions.

- Firms are *price takers*. There are so many firms in the industry that each one produces an insignificantly

Definition

Imperfect competition The collective name for monopolistic competition and oligopoly.

small proportion of total industry supply, and therefore has *no power whatsoever* to affect the price of the product. Hence, it faces a horizontal demand 'curve' at the market price: the price determined by the interaction of demand and supply in the whole market.

- There is complete *freedom of entry* into the industry for new firms. Existing firms are unable to stop new firms setting up in business. Setting up a business takes time, however. Freedom of entry, therefore, applies in the long run.
- All firms produce an *identical product*. (The product is 'homogeneous'.) There is therefore no branding or advertising, since there would be no point in the firm incurring this cost.
- Producers and consumers have *perfect knowledge* of the market. That is, producers are fully aware of prices, costs and market opportunities. Consumers are fully aware of price, quality and availability of the product.

These assumptions are very strict. Few, if any, industries in the real world meet these conditions. The foreign exchange market or certain agricultural markets, such as the market for fresh vegetables, are perhaps closest to perfect competition.

Nevertheless, despite the lack of real-world cases, the model of perfect competition plays a very important role in economic analysis and policy. Its major relevance is as an 'ideal type' against which to compare real-world markets.

Many argue that achieving perfect competition would bring a number of important advantages, such as keeping prices down to marginal cost and preventing firms from making supernormal profit over the long run. The model can thus be used as a benchmark against which to judge the shortcomings of particular industries. However, we shall also see that it has disadvantages, when compared with other market structures.

Pause for thought

It is sometimes claimed that the market for various stocks and shares is perfectly competitive, or nearly so. Take the case of the market for shares in a large company, such as Apple. Go through each of the four assumptions above and see if they apply in this case. (Don't be misled by the first assumption. The 'firm' in this case is not Apple itself, but rather the owners of the shares.)

The short-run equilibrium of the firm

In the **short run** we assume that the number of firms in the industry cannot be increased; there is simply not time for new firms to enter the market.

Figure 6.1 shows a short-run equilibrium for both industry and a firm under perfect competition. Both parts of the diagram have the same scale for the vertical axis. The horizontal axes have totally different scales, however. For example, if the horizontal axis for the firm were measured in, say, thousands of units, the horizontal axis for the whole industry might be measured in millions or tens of millions of units, depending on the number of firms in the industry.

TC 8 p 37

Let us examine the determination of price, output and profit in turn.

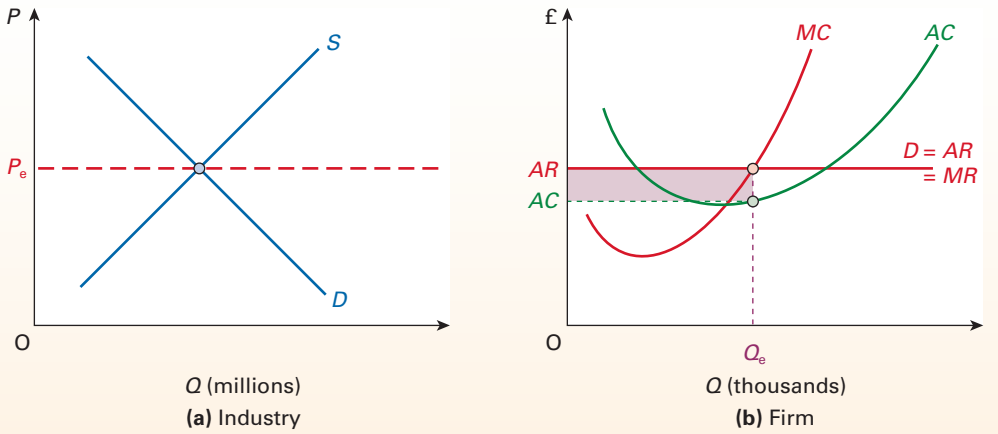
Price

The price is determined in the industry by the intersection of demand and supply. Being a price taker, the firm faces a horizontal demand (or average

Definition

Short run under perfect competition The period during which there is too little time for new firms to enter the industry.

Figure 6.1 Short-run equilibrium of industry and firm under perfect competition



revenue) 'curve' at this price. It can sell all it can produce at the market price (P_e), but nothing at a price above P_e, however, since competitors would be selling identical products at a lower price.

Output

The firm will maximise profit where marginal cost equals marginal revenue ($MR = MC$), at an output of Q_e. Note that, since the price is not affected by the firm's output, marginal revenue will equal price (see page 110 and Figure 5.4). Thus the firm's MR 'curve' and AR 'curve' (= demand 'curve') are the same horizontal straight line.

TC 2 p 9

Profit

If the average cost (AC) curve (which includes normal profit) dips below the average revenue (AR) 'curve', the firm will earn supernormal profit. Supernormal profit per unit at Q_e is the vertical difference between AR and AC at Q_e. Total supernormal profit is the shaded rectangle in Figure 6.1 (i.e. profit per unit times quantity sold).

The short-run supply curve

The *firm's* short-run supply curve will be its (short-run) marginal cost curve. But why? A supply curve shows how much will be supplied at each price: it relates quantity to price. The marginal cost curve relates quantity to marginal cost. But, under perfect competition, given that $P = MR$, and $MR = MC$, P must equal MC. Thus the supply curve and the MC curve will follow the same line.

For example, in Figure 6.2(b), if price were P_1, profits would be maximised at Q_1 where $P_1 = MC$. Thus point a is one point on the supply curve. At a price of P_2, Q_2 would be produced. Thus point b is another point on the supply curve, and so on.

KI 17 p 95

So, under perfect competition, the firm's supply curve depends entirely on production costs. This demonstrates why the firm's supply curve is upward sloping. Since marginal costs rise as output rises (due to diminishing marginal returns), a higher price will be necessary to induce the firm to increase its output.

Note that the firm will not produce at a price below AVC (see page 114 above). Thus the supply curve is only that portion of the MC curve above point e.

What will be the short-run supply curve of the whole *industry*? This is simply the sum of the short-run supply curves (and hence MC curves) of all the firms in the industry. Graphically this will be a *horizontal* sum, since it is *quantities* that are being added.

Pause for thought

Will the industry supply be zero below a price of P_5 in Figure 6.2?

The long-run equilibrium of the firm

In the **long run,** if typical firms are making supernormal profits, new firms will be attracted into the industry. Likewise, if existing firms can make supernormal profits by increasing the scale of their operations, they will do so, since all factors of production are variable in the long run.

TC 6 p 24

The effect of the entry of new firms and/or the expansion of existing firms is to increase industry supply. This is illustrated in Figure 6.3. At a price of P_1 supernormal profits are earned. The industry supply curve will thus shift to the right as new firms enter. This, in turn, leads to a fall in the market price.

KI 8 p 24

Definition

The long run under perfect competition The period of time that is long enough for new firms to enter the industry.

Figure 6.2 Deriving the short-run supply curve

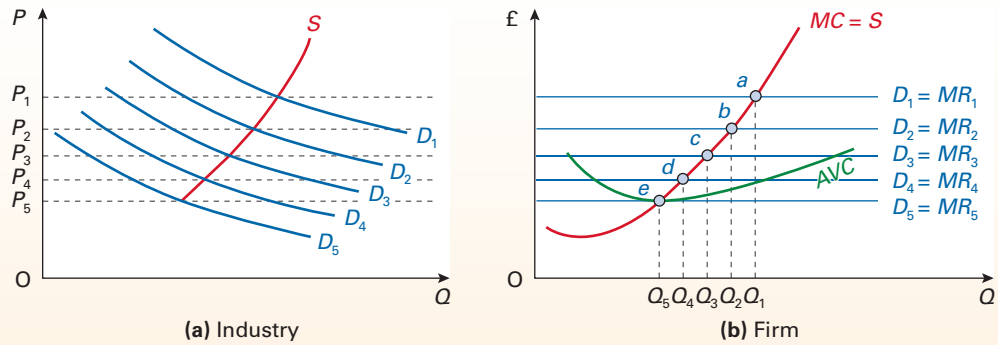
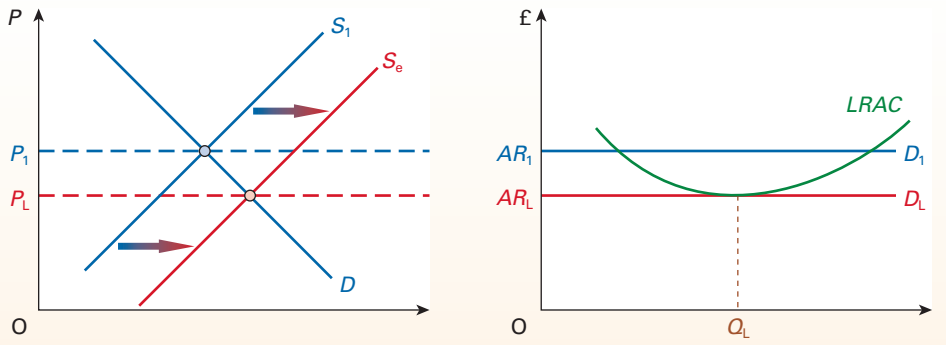


Figure 6.3 Long-run equilibrium under perfect competition



Supply will go on increasing, and the price will go on falling, until firms are making only normal profits. This will be when price has fallen to the point where the demand 'curve' for the firm just touches the bottom of its long-run average cost curve. Q_L is thus the long-run equilibrium output of the firm, with P_L the long-run equilibrium price.

TC 8 p 37

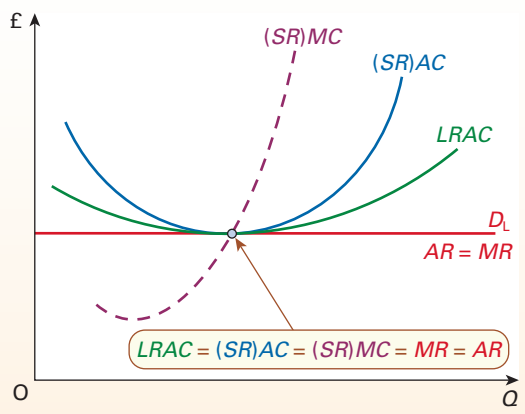
Since the *LRAC* curve is tangential to (i.e. just touching) all possible short-run *AC* curves (see Section 5.2), the full long-run equilibrium will be as shown in Figure 6.4 where:

$$LRAC = AC = MC = MR = AR$$

The incompatibility of perfect competition and substantial economies of scale

Why is perfect competition so rare in the real world – if it even exists at all? One important reason for this has to do with *economies of scale*.

Figure 6.4 Long-run equilibrium of the firm under perfect competition



In many industries, firms may have to be quite large if they are to experience the full potential economies of scale. But perfect competition requires there to be *many* firms. Firms must therefore be small under perfect competition: too small in most cases for economies of scale.

Once a firm expands sufficiently to achieve economies of scale, it will usually gain market power. It will be able to undercut the prices of smaller firms, which will thus be driven out of business. Perfect competition is destroyed.

KI 19 p 122

Perfect competition could exist in any industry, therefore, only if there were no (or virtually no) economies of scale.

Is perfect competition good for consumers?

Generally it is argued that perfect competition is a 'good thing', and that the more perfect an industry becomes the better. We explore the arguments in Section 6.3 (pages 132–3) after we have looked at monopoly, but at this stage the main points in favour of perfect competition can be identified.

Price equals marginal cost ($P = MC$). Why is this desirable? To answer this, consider what would happen if they were not equal.

If price were greater than marginal cost ($P > MC$), this would mean that consumers were putting a higher value (P) on the production of extra units than they cost to produce (MC). Therefore *more* ought to be produced. If price were less than marginal cost ($P < MC$), consumers would be putting a lower value on extra units than they cost to produce. Therefore less ought to be produced. When they are equal, therefore, production levels are just right. But, as we shall see later, it is only under perfect competition that $MC = P$.

This idea of producing just the right amount of the product is referred to as **allocative efficiency**.

Least-cost output. Long-run equilibrium is at the bottom of the firm's long-run *AC* curve. That is, for any *given* technology, the firm in the long run will produce at the least-cost output.

Keeps prices at a minimum. The combination of (long-run) production being at minimum average cost and the firm making only normal profit keeps prices at a minimum.

Productive efficiency. Perfect competition is a case of 'survival of the fittest'. Inefficient firms will be driven out of business, since they will not be able to make even normal profits. This encourages firms to be as efficient as possible and, where possible, to invest in new improved technology. Therefore, firms choose the cheapest possible way of producing any level of output and to do so allocate the factors of production optimally. This is the idea of *productive efficiency* introduced in Chapter 5 (see page 106).

In general, it can be argued that perfectly competitive markets result in economic efficiency.

KEY IDEA 20

Economic efficiency **is achieved when each good is produced at the minimum cost and where consumers get maximum benefit from their income.**

Recap

1. The assumptions of perfect competition are: a very large number of firms, complete freedom of entry, a homogeneous product and perfect knowledge of the good and its market by both producers and consumers.
2. In the short run, there is not time for new firms to enter the market, and thus supernormal profits can persist. In the long run, however, any supernormal profits will be competed away by the entry of new firms.
3. The short-run equilibrium for the firm is where the price, as determined by demand and supply in the market, is equal to marginal cost. At this output the firm will be maximising profit.
4. The long-run equilibrium is where the market price is just equal to firms' long-run average cost.
5. There can be no substantial economies of scale to be gained in a perfectly competitive industry. If there were, the industry would cease to be perfectly competitive as the large, low-cost firms drove the small, high-cost ones out of business.

6.3 MONOPOLY

What happens when there is only one firm in the market? Do we as consumers suffer?

What is a monopoly?

This may seem a strange question because the answer appears obvious. A monopoly exists when there is only one firm in the industry.

But whether an industry can be classed as a monopoly is not always clear. It depends how narrowly the industry is defined. For example, a confectionary company may have a monopoly on certain chocolate bars, but it does not have a monopoly on chocolate in general. A pharmaceutical company may have a monopoly of a certain drug, but there may be alternative drugs for treating a particular illness.

KI 19 p 122

To some extent, the boundaries of an industry are arbitrary. What is more important for a firm is the amount of monopoly *power* it has, and that depends on the closeness of substitutes produced by rival industries. A train company may have a monopoly over railway journeys between two towns, but it faces competition in transport from cars, coaches and sometimes planes.

Barriers to entry

For a firm to maintain its monopoly position, there must be **barriers to entry** that make it difficult for new firms to enter the market. As we shall see, barriers also exist under oligopoly, but in the case of monopoly they must be high enough to block the entry of new firms. Barriers can take various forms.

Economies of scale. If the monopolist's costs go on falling significantly up to the output that satisfies the

Definitions

Allocative efficiency A situation where the current combination of goods produced and sold gives the maximum satisfaction for each consumer at their current levels of income.

Barriers to entry Anything that prevents or impedes the entry of firms into an industry and thereby limits the amount of competition faced by existing firms.

BOX 6.1 E-COMMERCE

Has technology shifted market power?

Despite the growth in big business creating more concentrated markets dominated by large producers, forces have been working to bring more competition to markets. One of these forces is e-commerce and COVID-19 accelerated this change.

In this box, we consider the impact of e-commerce on market structures and power.

The rise of e-commerce

E-commerce, or e-shopping, has grown rapidly in the past 15 years and shows no sign of slowing down. The chart shows the rise in online retail sales as a proportion of all retail sales in the UK between 2007 and 2022. Note that the proportion of Internet sales rises each year in the run-up to Christmas as many people buy gifts online.

During the COVID-19 pandemic, businesses offering online shopping or those that invested in it quickly gained over those which relied on people coming into their stores. The pandemic showed the importance of online shopping capability and many firms have since invested in this or improved their offering.

Moving markets back towards perfect competition?

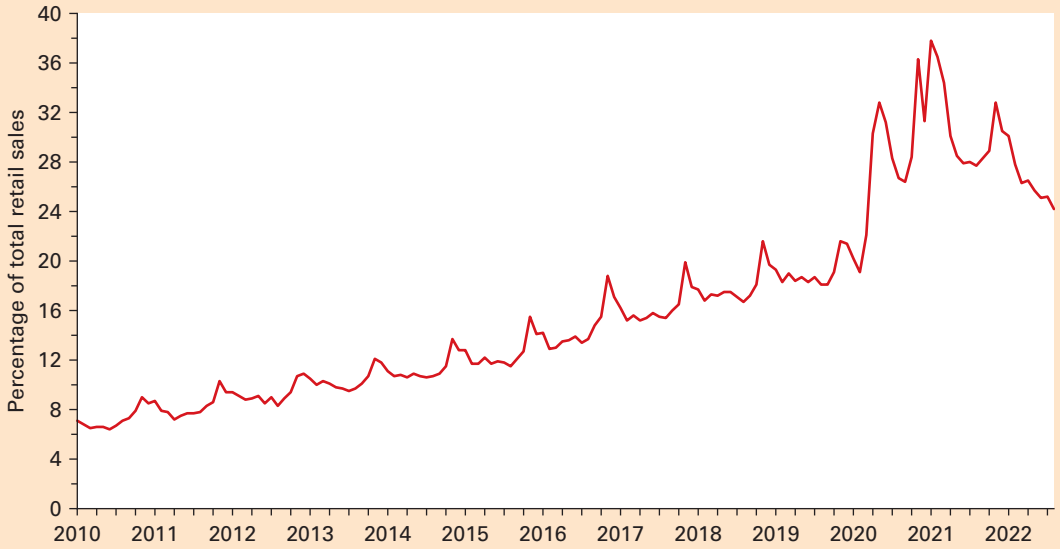
Let us reconsider three of the assumptions of perfect competition and how they apply to e-commerce.

A large number of buyers and sellers. The growth of e-commerce has led to many new firms starting up in business. The majority of these are small companies that often sell directly to consumers via their own websites or use online market places such as eBay or Amazon. As of the final quarter of 2021, eBay had 147 million buyers, 17 million sellers and 1.5 billion live listings, while Amazon had 2 million third-party sellers (accounting for approximately 50 per cent of sales volume), reaching 310 million customers, with 200 million Amazon Prime members.

The global reach of the Internet increases the potential number of buyers and sellers that can trade with one another. Firms must now compete with others across the world, as consumers have access to the global marketplace. They must keep an eye on the prices and products of competitors worldwide and be aware of the continual emergence of new smaller businesses.

Freedom of entry. Internet companies often have lower start-up costs than their conventional rivals. Their premises are generally much smaller, with no 'shop-front' costs and lower levels of stockholding; in fact, many of these businesses are initially operated from their owners' homes. and garages with little more required than a computer and good WiFi connection. An e-commerce website for the business can be set up in a matter of hours. Marketing costs will also be lower if the new entrant's website can easily be located by a consumer using a search engine.

Value of Internet retail sales as a percentage of total retail sales



Source: Based on series J4MC from *Time Series Data* (National Statistics)

whole market, the industry may not be able to support more than one producer. This case is known as **natural monopoly**. It is particularly likely if the market is small and/or the industry has relatively high capital/infrastructure costs (i.e. fixed costs) and relatively low marginal costs.

Definition

Natural monopoly A situation where long-run average costs would be lower if an industry were under monopoly than if it were shared between two or more competitors.

CASE STUDIES & APPLICATIONS

Internet companies are often smaller and more specialist, relying on Internet 'outsourcing' (buying parts, equipment and other supplies through the Internet), rather than making everything themselves. This can allow them to economise on stockholding. They are also more likely to use delivery firms rather than having their own transport fleet, helping to keep costs down and making their businesses more viable.

The market has never been more open to new competitors and this has spill-over effects on existing firms, which are forced to become more efficient. Demand curves have become more elastic, more innovative products are available and markets have become more price competitive.

Perfect knowledge. The Internet has significantly reduced some of the transaction costs of using a market. Consumers can easily compare the prices and other features of the goods they are interested in purchasing by using search engines and comparison sites. It is common to see people in shops (physical shops in this case) browsing competitors' prices on their mobile phones. This places the High Street retailer under intense competitive pressure. Interestingly, several of the top online retailers in the UK – Tesco, Argos, John Lewis, Next – also have physical shops. Success for some firms now means having both physical and online stores.

In the final quarter of 2020, both Debenhams and Arcadia went into administration. Boohoo, the online fashion retailer, purchased the Debenhams brand in January 2021 but not the department stores. Another online retailer, Asos, purchased the Topshop, Topman, Miss Selfridge and HIIT brand names from Arcadia in February 2021. Once again, the deal did not include the stores.

Although the competitive pressures seem to have increased in 'B2C' (business-to-consumers) e-commerce, the impact may be even greater in 'B2B' (business-to-business) e-commerce. Many firms are constantly searching for cheaper sources of supply, and the Internet provides a cheap and easy means of conducting such searches.

Increasing market power?

There are, however, some concerns that e-commerce could actually reduce competition and result in the growth of more firms with substantial market power.

Greater price transparency might actually result in less competition. Before the Internet, firms could often gain a lasting advantage by cutting price. However, firms can now observe and respond instantly if their rivals cut price, meaning any advantage is removed quickly. This may remove the incentive for firms to cut prices and even encourage collusion to charge higher prices.

Although comparison websites increase price transparency, they can also result in consumers paying higher prices. Many of these websites make considerable profits. They earn revenue by charging a fee every time a customer is referred to a listed firm's website via the price comparison website. These fees add to costs and could result in firms charging higher prices.

While marketing costs for start-ups in e-commerce can be lower than those in more traditional retailing, new firms might have to spend considerable amounts of money on marketing to increase consumers' awareness of their brands and websites. Larger firms, such as Amazon, also benefit from the lower costs that technology and online marketplaces bring. It is also these more established firms which are likely to be listed towards the top of a search engine's results page and thus more likely to be clicked on by customers. Also, if users build up a familiarity and knowledge of using a particular website, such as the Amazon site, it might create switching costs. Consumers are then less likely to visit other rival websites and so this reduces competition.

There may also be significant economies of scale in logistics: i.e. the storage, packaging and shipping of the products to the consumer. Amazon has invested heavily in automating its distribution centres using Kiva robots. This type of capital investment will only reduce a firm's average costs if it sells a large volume of products.

1. *Why may the Internet work better for replacement buys than for new purchases?*
2. *Give three examples of products that are particularly suitable for selling over the Internet and three that are not. Explain your answer.*
3. *As Amazon has grown in size, it has acquired substantial monopoly power. What are the barriers to entry for other companies wishing to act as a marketplace for B2C and B2B business?*

Using the Eurostat database construct a time series chart showing the trend in the percentage of individuals who have made an online purchase in the past three months in the UK, France, Germany, Ireland, and Spain. Write a short summary of your findings.

One real-world example is the network of pipelines that supply gas to homes and businesses. If two competing firms each built a national network of pipes it might be difficult for them both to make a profit as they would share the customers but each have the same infrastructure costs. A monopoly that supplies all customers could make a profit as it would have much lower average total costs from supplying the whole market.

Even if a market could support more than one firm, a new entrant is unlikely to be able to start up on a very large scale. Thus the monopolist that is already experiencing economies of scale can charge a price below the cost of the new entrant and drive it out of business.

If, however, the new entrant is a firm already established in another industry, it may be able to survive this competition. For example, Amazon entered the UK online grocery market in 2016.

Absolute cost advantages. If a monopolist has an absolute cost advantage, its average cost curve will be below that of any potential entrants at all levels of output. What might give a monopolist such a cost advantage?

- *More favourable access to or control over key inputs.* In some markets the monopolist might be able to obtain access to important factor inputs on more favourable terms for a certain period of time. For example, if there was a supplier that provided a much higher quality of a factor input than its rivals, the monopoly could either sign a long-term exclusive contract with this firm or take ownership via a merger or takeover. For example, in 2012 Amazon purchased Kiva Systems. This company was the leading supplier of robotics for a number of warehouse operators and retailers. After the takeover, Kiva only supplied Amazon and was renamed Amazon Robotics in 2015.

  In more extreme cases the monopolist may be able to gain complete control if there is only one supplier of that input. For many years the De Beers company owned both the majority of the world's diamond mines and the major distribution system.
- *Superior technology.* The monopolist may have access to superior technology that is difficult for rival firms either to copy or to imitate. For many years, Google's search ranking algorithm helped it to provide a results page that many people found more useful than those of its rivals.
- *More efficient production methods.* Through years of experience of running the business, an established monopoly might have learnt the most efficient way of organising the production of its good or service. Much of this knowledge is tacit. It is developed and refined through a process of trial and error and cannot be written down in a way which could be easily understood by others. The new entrant would have to go through the same learning experience over a number of years before it could operate on the same cost curve as the monopolist.
- *Economies of scope.* A firm that produces a *range* of products is also likely to experience a lower average cost of production. For example, a large pharmaceutical company producing a range of drugs and toiletries can use shared research, marketing, storage and transport facilities across its range of products. These lower costs make it difficult for a new single-product entrant to the market, since the large firm will be able to undercut its price and drive it out of the market.

Pause for thought

Under what circumstances might a new entrant succeed in the market for a product, despite existing firms benefiting from economies of scale?

Switching costs. Sometimes, if customers are considering whether or not to buy a product from a different firm, they may decide against it because of the additional costs involved. These are called **switching costs** and some examples include:

- *Searching costs.* The more difficult it is for the consumer to find and compare the price and quality of goods/services offered by alternative suppliers, the greater the switching costs. Some firms have been accused of deliberately making it more difficult for consumers to make these comparisons. For example, a report into the retail banking market by the Competition and Markets Authority in 2016[1] found low switching rates among current account holders, because they found it almost impossible to compare products.
- *Contractual costs.* In some markets customers must sign a contract which stipulates that they purchase the good or service from the same supplier for a certain period of time: e.g. energy, mobile phones, broadband. A termination fee has to be paid if the customer wants to switch to a different supplier before the end of the contract period. Some firms also provide incentives for repeat purchases: e.g. loyalty cards, frequent flyer programmes.
- *Learning costs.* These may occur if the consumer invests time and effort in learning how to use a product or service. The switching costs increase as this knowledge becomes more specific to the brand/product supplied by a particular firm. For example, a consumer might have spent a considerable amount of time learning how to use applications with the iOS operating system on an iPhone. This may deter them from switching to a smartphone that uses Google's Android operating system.
- *Product uncertainty costs.* Consumers might not fully discover either the quality or how much they like a good until after they have purchased and used it for some time. This might make them reluctant to

KI 15 p 80

Definition

Switching costs The costs to a consumer of switching to an alternative supplier.

[1] *Retail banking market investigation: Final report*, Competition and Markets Authority (9/8/16) (See section 14).

change supplier once they have found and experienced a particular brand or product they like.

- *Compatibility costs.* Some products have two elements to them. One part is more durable, while the other needs replacing more regularly. Once customers have purchased the more durable element from a supplier, they are 'locked in' to purchasing the non-durable part from the same supplier for compatibility reasons. Examples include razor handles and razor blades, coffee machines and coffee pods, printers and ink cartridges.
- *Network economies.* Some goods or services have very large switching costs because of the existence of network externalities. A **network externality** exists when consumers' valuation of a good is influenced by the number of other people who also use the same product. For example, the benefits of having a mobile phone increase with the number of other people who also have one.

In some cases, a consumer's valuation will depend on the extent to which other people use one particular brand. For a specific social media website, such as Facebook, its success depends on lots of people using the same website. Buyers and sellers are willing to pay higher fees to use Amazon and eBay as an online marketplace because so many other buyers and sellers use the same websites.

When a good or service has significant network externalities it makes it difficult for a new entrant. Even if it produces a far superior and/or cheaper version of a product, it is difficult to get people to switch because they are unwilling to give up the network benefits associated with their current supplier. Other examples of products with network externalities include Microsoft Windows® (see Case Study 6.4 on the student website), Adobe Acrobat® (for PDF files) and airlines operating interconnecting routes (see Case Study 6.6 on the student website).

Product differentiation and brand loyalty. If a firm produces a clearly differentiated product, where the consumer associates the product with the brand, it will be very difficult for a new firm to break into that market.

In 1908 James Spengler invented, and patented, the electric vacuum cleaner. Later that year he sold the patent to his cousin's husband, William Hoover, who set about putting mass-production in place. Decades after their legal monopoly (see below) ran out, people still associate vacuum-cleaning with Hoover® and many of us would say that we are going to 'Hoover the carpet', despite using a Dyson, or other machine. When looking for some information by using an Internet search engine people often say 'they googled it'.

Other examples of strong brand image include Guinness®, Kellogg's Cornflakes®, Coca-Cola, Nescafé® and Sellotape® (or Scotch Tape® in the USA). In many cases, strong brand presence would not be enough to *block* entry, but it might well reinforce other barriers.

More favourable or complete control over access to customers. If a firm can gain more favourable access or control over the best outlets through which the product is sold, this can hinder the ability of new entrants to gain access to potential customers. For example, approximately 50 per cent of public houses (pubs) in the UK operate on tenancy contracts known as the 'tied lease model'. This is effectively an exclusive supply contract which means that landlords of such pubs have to purchase almost all of their beverages from the pub company (e.g. Enterprise Inns, Punch Taverns and JD Wetherspoon) that owns the pub.

Legal protection. The firm's monopoly position may be protected by patents on essential processes, by copyright, by various forms of licensing (allowing, say, only one firm to operate in a particular area) and by tariffs (i.e. customs duties) and other trade restrictions to keep out foreign competitors. Examples of monopolies protected by patents include most new medicines developed by pharmaceutical companies (e.g. anti-AIDS drugs), Microsoft's Windows operating systems, and agrochemical companies, such as Monsanto, with various genetically modified plant varieties and pesticides.

Mergers and takeovers. The monopolist can put in a takeover bid for any new entrant. The sheer threat of takeovers may discourage new entrants.

Aggressive tactics. An established monopolist can probably sustain losses for longer than a new entrant. Thus it can start a price war, mount massive advertising campaigns, offer an attractive after-sales service, introduce new brands to compete with new entrants, and so on.

Intimidation. The monopolist may resort to various forms of harassment, legal or illegal, to drive a new entrant out of business.

Equilibrium price and output

Since there is, by definition, only one firm in the industry, the firm's demand curve is also the industry demand curve.

Definition

Network externalities or network economies The benefits a consumer obtains from consuming a good/service increase with the number of other people who use the same good/service.

Compared with other market structures, demand under monopoly tends to be less elastic at each price. TC 10 p 37 The monopolist can raise its price and consumers have no alternative firm to turn to within the industry. They either pay the higher price or go without the good altogether.

Unlike the firm under perfect competition, the monopoly firm is thus a 'price maker'. It can choose what price to charge. Nevertheless, it is still constrained by its demand curve. A rise in price will reduce the quantity demanded.

As with firms in other market structures, a monopolist will maximise profit where $MR = MC$. TC 2 p 9 In Figure 6.5 profit is maximised at Q_m. The supernormal profit obtained is shown by the shaded area.

These profits will tend to be larger the less elastic is the demand curve (and hence the steeper is the *MR* curve), and thus the bigger is the gap between *MR* and price (*AR*). The actual elasticity will depend on whether reasonably close substitutes are available in *other* industries. The demand for a rail service will be much less elastic (and the potential for profit greater) if there is no bus service to the same destination.

Since there are barriers to the entry of new firms, a monopolist's supernormal profits will not be competed away in the long run. The only difference, therefore, between short-run and long-run equilibrium is that in the long run the firm will produce where $MR =$ *long-run* MC.

Monopoly versus perfect competition: which best serves the public interest?

Because it faces a different type of market environment, the monopolist will produce a quite different output and at a quite different price from a perfectly competitive industry. Let us compare the two.

Short-run price and output. Figure 6.6 compares the profit-maximising position for an industry under monopoly with that under perfect competition. Note that we are comparing the monopoly with the whole *industry* under perfect competition. That way we can assume, for sake of comparison, that they both face the same demand curve. We also assume for the moment that they both face the same cost curves.

The monopolist will produce Q_1 at a price of P_1. KI 19 p 122 This is where $MC = MR$. If the same industry were under perfect competition, however, it would produce at Q_2 and P_2 – a higher output and a lower price. But why? The reason is that for each of the firms in the industry – and it is at this level that the decisions are made – marginal revenue is the same as price. Remember that the *firm* under perfect competition faces a perfectly elastic demand (*AR*) curve, which also equals *MR* (see Figure 6.1). Thus producing where $MC = MR$ also means producing where $MC = P$. When *all* firms under perfect competition do this, price and quantity in the *industry* will be given by P_2 and Q_2 in Figure 6.6.

In the short run, therefore, it would seem (other things being equal) that perfect competition better serves the consumer's interest than does monopoly.

Long-run price and output. Under perfect competition, freedom of entry eliminates supernormal profit and forces firms to produce at the bottom of their *LRAC* curve. The effect, therefore, is to keep long-run prices down. Under monopoly, however, barriers to entry allow profits to remain supernormal in the long run. The monopolist is not forced to operate at the bottom

Figure 6.5 Profit maximising under monopoly

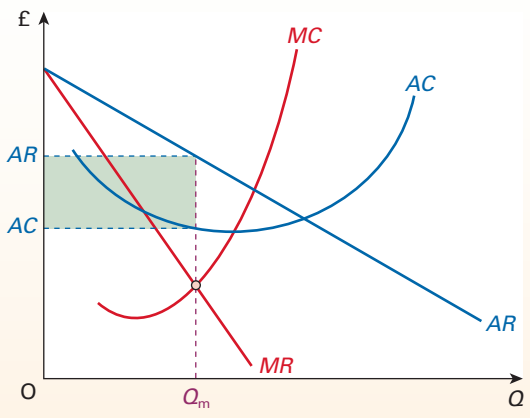
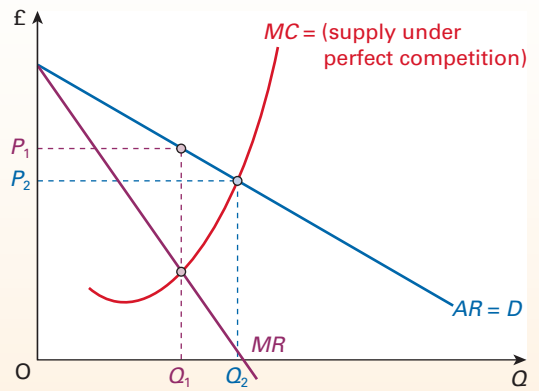


Figure 6.6 Equilibrium of the industry under perfect competition and monopoly: with the same *MC* curve



of the *AC* curve. Thus, other things being equal, long-run prices will tend to be higher, and hence output lower, under monopoly.

KI 19 p 122

Pause for thought

If the shares in a monopoly (such as a water company) were very widely distributed among the population, would the shareholders necessarily want the firm to use its monopoly power to make larger profits?

Thus, again, it would *seem* that perfect competition better serves the consumer's interests. But this assumes that the cost curves are the *same* under both perfect competition and monopoly. Let us, therefore, turn to costs.

Costs under monopoly. The sheer survival of a firm in the long run under perfect competition requires that it uses the most efficient known technique, and develops new techniques wherever possible. The monopolist, however, sheltered by barriers to entry, can still make large profits even if it is not using the most efficient technique. It has less incentive, therefore, to be efficient. For this reason, costs may be *higher* under monopoly (another criticism of monopoly).

TC 6 p 24

On the other hand, the monopoly may be able to achieve substantial economies of scale due to larger plant, centralised administration and the avoidance of unnecessary duplication (e.g. a monopoly water company would eliminate the need for several sets of rival water mains under each street). If this results in an *MC* curve substantially below that of the same industry under perfect competition, the monopoly may even produce a *higher* output at a *lower* price.

Another reason why a monopolist may operate with lower costs is that it can use part of its supernormal profits for research and development and investment. It may not have the same *incentive* to become efficient as the perfectly competitive firm which is fighting for survival, but it may have a much greater *ability* to become efficient than has the small firm with limited funds.

Although a monopoly faces no competition in the goods market, it may face an alternative form of competition in financial markets. A monopoly, with potentially low costs, which is currently run inefficiently, is likely to be subject to a takeover bid from another company. This **competition for corporate control** may thus force the monopoly to be efficient in order to prevent being taken over.

TC 6 p 24

Innovation and new products. The promise of supernormal profits, protected perhaps by patents, may encourage the development of new (monopoly) industries producing new products. It is this chance of making monopoly profits that encourages many people to take the risks of going into business.

KI 15 p 80

Potential competition or potential monopoly? The theory of contestable markets

Potential competition

The theory of contestable markets argues that what is crucial in determining price and output is not whether an industry is *actually* a monopoly or competitive, but whether there is the real *threat* of competition.

If a monopoly is protected by high barriers to entry – say, it controls the supply of key raw materials – then it will be able to make supernormal profits with no fear of competition.

If, however, another firm *could* take over from it with little difficulty, it will behave much more like a competitive firm. The threat of competition has a similar effect to actual competition.

As an example, consider a catering company that is given permission by a university to run its cafés and coffee bars. The catering company has a monopoly over the supply of food and drinks to the students at the university, assuming there are no other eating places nearby. If, however, it starts charging high prices or providing a poor service, the university could offer the running of the cafés to an alternative catering company. This threat may force the original catering company to charge 'reasonable' prices and offer a good service.

Perfectly contestable markets

A market is **perfectly contestable** when the costs of entry and exit by potential rivals are zero, and when such entry can be made very rapidly. In such cases, the moment the possibility of earning supernormal profits occurs, new firms will enter, thus driving profits down to a normal level.

TC 7 p 26

The sheer threat of this happening, so the theory goes, will ensure that the firm already in the market will (a) keep its prices down, so that it just makes normal profits, and (b) produce as efficiently as possible, taking advantage of any economies of scale and any new technology. If the existing firm did not do this, entry would take place, and potential competition would become actual competition.

Definitions

Competition for corporate control The competition for the control of companies through takeovers.

Perfectly contestable market A market where there is free and costless entry and exit.

BOX 6.2 BREAKING THE MONOPOLY ON LIVE PREMIER LEAGUE FOOTBALL CASE STUDIES & APPLICATIONS

The sky is the limit for the English Premier League

The early days of Sky

In the 1992/3 football season, the FA Premier League (EPL) was formed. Live league football moved off free-to-air television, with the hope of higher payments from TV companies for the rights to show live matches. The first contract to acquire the exclusive EPL broadcasting rights for the UK and Republic of Ireland was worth £191m over five seasons. Sky won the contract, meaning anyone wanting to watch live football matches on television had to sign up to Sky, buying a basic package and the Sky sports channels; Sky thus had a monopoly on football broadcasts.

Attempts to reduce monopoly power

Both the Premier League and Sky's coverage were hugely successful with viewers and advertisers, reflected in the price paid for subsequent packages. The four seasons from 1997 cost Sky £670m, while BSkyB (Sky's new name) paid over £1bn in 2003 for exclusive rights for three seasons.

Over this period, the European Commission expressed increasing concern about the extent of Sky's monopoly. It started legal proceedings in 2002, filing a statement of objections, but was thwarted when the League agreed a new contract with Sky before ironing out an EC-approved deal.

Sky agreed to sub-license eight 'top quality Premier League matches' each season to another broadcaster in order to win European approval. The Commission trumpeted this pledge as meaning that, for the first time, free-to-air television would have a realistic opportunity to show live Premier League matches. These hopes were dashed when no rival broadcaster met Sky's asking price.

Auctioning the TV rights to the EPL

In 2005 the European Commission announced that Sky's monopoly would be broken. From 2007, the rights for the next three-season period would be sold in six 'balanced' packages of 23 games per season, with no broadcaster allowed more than five packages. The Commission claimed the deal would give fans 'greater choice and better value'.

However, concern was expressed about the impact on incomes of Premiership clubs. Some commentators expected a more competitive process would generate a similar amount of total income. Others, however, suggested that Sky's 2003 payment included a premium for the guarantee that it would be the sole broadcaster. With bidding now being competitive, this would result in a fall in the revenues paid to clubs.

In May 2006 the bidding process for the rights for 2007–10 was completed. Sky won four of the six packages, showing 92 live Premiership matches per season. Setanta, an Irish-based satellite broadcaster, won the remaining two packages and showed 46 games per season. Together, they paid £1.7bn. The same process was undertaken in 2009 for the 2010/11 to 2013/14 seasons, generating just under £1.8bn.

In 2012 Setanta folded and BT entered the market. Despite there still being two bidders, the total amount paid via the auctions for domestic EPL TV rights increased, with BT and Sky jointly paying £3bn. The 2015 auction was acutely competitive, as both companies saw the securing of the rights to screen live EPL matches as a means by which they could sell highly profitable broadband and phone services bundled with their respective TV offering. The result was Sky paying £4.17bn for five of the auctioned packages (£11.047m per game) with BT paying £960m for two packages (£7.619m per game), generating £5.13bn in total.

Prior to the 2018 auction, Sky and BT struck a deal to carry each other's content on their platforms. For some, this marked the end to an aggressive battle for dominance of the UK pay-TV market and this was reflected in the prices paid for live EPL matches from 2019/20 to 2022/23. The initial announcement revealed that Sky and BT had paid £4.464bn for five of the seven packages, with Sky paying £3.579bn (£9.3m per game) and BT £885m (£9.22m per game). The final two packages, each with 20 games, were not initially allocated, failing to meet the reservation price.

The end of the auction?

Ofcom wanted more live matches on TV, but this meant multiple live matches being shown at the same time. The final packages took over four months to sell. They were expected to make over £100m, but BT paid only £30m for one package (£1.5m per match), while a new entrant paid an undisclosed (and probably lower) sum for the other. The good news: the new entrant was Amazon.

At the time, the revenue from 2018 was 'disappointing', but since then, with COVID-19 impacting revenues, the amount generated is now seen as acceptable. There was little appetite amongst the TV providers to enter a bidding war for the 2022/23 to 2025/26 seasons of EPL games, particularly with little chance of new bidders. To avoid an auction wiping off hundreds of millions of pounds, EPL officials negotiated with its existing broadcast partners and the UK government to extend their deals, worth £4.8bn. These included an extra £100m of 'trickle-down funding' over four years to support parts of the football community particularly affected by COVID-19. This is likely to prove very helpful for teams in lower divisions, given the estimated £2bn that clubs lost through ticket sales and rebates to broadcasters over cancelled matches.

In August 2021, the UK government agreed to the deal, granting an exclusion under the Competition Act to avoid the normal tendering process, as a 'temporary measure' because of the pandemic. How rights to UK broadcasters will be decided post-2025 remains to be seen.

1. *What other examples of monopoly power exist in football? Could this power be reduced?*
2. *Assess the impact of the pay broadcasters' emergence since the establishment of the English Premier League on (a) football fans, (b) other viewers.*
3. *Why might a company like Amazon have more chance of successfully entering the pay-TV market than alternative telecommunications companies?*

Research the allocation of rights to broadcast live top-flight football in another European league. How does the approach differ from the EPL? Assess its impact on the revenues raised, TV viewers and football fans.

Contestable markets and natural monopolies

So why in such cases are the markets not *actually* perfectly competitive? Why do they remain monopolies?

The most likely reason has to do with economies of scale and the size of the market. To operate on a minimum efficient scale, the firm may have to be so large relative to the market that there is only room for one such firm in the industry. If a new firm does come into the market, then one or other of the two firms will not survive the competition. The market is simply not big enough for both of them.

If, however, there are no entry or exit costs, new firms will be perfectly willing to enter even though there is only room for one firm, provided they believe that they are more efficient than the existing firm. The existing firm, knowing this, will be forced to produce as efficiently as possible and with only normal profit.

The importance of costless exit

Setting up in a new business usually involves large expenditures on plant and machinery. Once this money has been committed, it becomes fixed costs. If these fixed costs are no higher than those of the existing firm, then the new firm could win the battle. But, of course, there is always the risk that it might lose.

But does losing the battle really matter? Can the firm not simply move to another market?

It does matter if there are substantial costs of exit. This will be the case if the capital equipment cannot be transferred to other uses – for example, a new blast furnace constructed by a new rival steel company. In this case, these fixed costs are *sunk costs* in production (see page 96). The losing firm is left with capital equipment it cannot use. The firm may therefore be put off entering in the first place. The market is not perfectly contestable, and the established firm can make supernormal profit.

If, however, the capital equipment can be transferred, the exit costs will be zero (or at least very low), and new firms will be more willing to take the risks of entry. For example, a rival coach company may open up a service on a route previously operated by only one company, and where there is still only room for one operator. If the new firm loses the resulting battle, it can still use the coaches it has purchased. It simply uses them for a different route. The cost of the coaches is not a sunk cost.

Costless exit, therefore, encourages firms to enter an industry, knowing that, if unsuccessful, they can always transfer their capital elsewhere.

The lower the exit costs, the more contestable the market. This implies that firms already established in other similar markets may provide more effective competition against monopolists, since they can simply transfer capital from one market to another. For example, studies of airlines in the USA show that entry to a particular route may be much easier for an established airline, which can simply transfer planes from one route to another.

Contestability and the public interest

The more contestable the market, the more a monopoly will be forced to act like a firm under perfect competition. If, therefore, a monopoly operates in a perfectly contestable market, it might bring the 'best of both worlds' for the consumer. Not only will it be

KI 20 p 127

Recap

1. A monopoly is where there is only one firm in an industry. In practice, it is difficult to determine where a monopoly exists because it depends on how narrowly an industry is defined.
2. Barriers to the entry of new firms will normally be necessary to protect a monopoly from competition. Such barriers include economies of scale (making the firm a natural monopoly or at least giving it a cost advantage over new (small) competitors), control over supplies of inputs or over outlets, patents or copyright, and tactics to eliminate competition (such as takeovers or aggressive advertising).
3. Profits for the monopolist will be maximised (as for other firms) where $MC = MR$.
4. If demand and cost curves are the same in a monopoly and a perfectly competitive industry, the monopoly will produce a lower output and at a higher price than the perfectly competitive industry.
5. On the other hand, any economies of scale will, in part, be passed on to consumers in lower prices, and the monopolist's high profits may be used for research and development and investment, which in turn may lead to better products at possibly lower prices.
6. Potential competition may be as important as actual competition in determining a firm's price and output strategy.
7. The threat of this competition is greater the lower the entry and exit costs are to and from the industry. If the entry and exit costs are zero, the market is said to be perfectly contestable. Under such circumstances an existing monopolist will be forced to keep its profits down to the normal level if it is to resist the entry of new firms. Exit costs will be lower, the lower the sunk costs of the firm are.

able to achieve low costs through economies of scale, but also the potential competition will keep profits and hence prices down.

For this reason, the theory has been seized on by politicians on the political right to justify a policy of *laissez-faire* (non-intervention) and deregulation (e.g. coach and air routes). They argue that the theory vindicates the free market. However, in responding to this, others argue that few markets are *perfectly* contestable. If entry and exit are not costless, a monopoly can still make supernormal profits in the long run.

Furthermore, there are other possible failings of the market beside monopoly power (e.g. inequality and pollution). These failings are examined in Chapters 7 and 8.

Pause for thought

Think of two examples of highly contestable monopolies (or oligopolies). How well is the public interest served?

6.4 MONOPOLISTIC COMPETITION

What happens if there are quite a lot of firms competing, but each firm tries to attract us to its particular product or service?

Very few markets in practice can be classified as perfectly competitive or as a pure monopoly. The vast majority of firms do compete with other firms, often quite aggressively, and yet they are not price takers: they do have some degree of market power. Most markets, therefore, lie between the two extremes of monopoly and perfect competition, in the realm of 'imperfect competition'. As we saw in Section 6.1, there are two types of imperfect competition: namely, monopolistic competition and oligopoly.

Monopolistic competition is nearer to the competitive end of the spectrum. It can best be understood as a situation where there are a lot of firms competing but where each firm does nevertheless have some degree of market power (hence the term 'monopolistic' competition): each firm has some discretion as to what price to charge for its products.

Assumptions

TC 3 p 10

- There are *quite a large number of firms*. As a result each firm has only a small share of the market, and therefore its actions are unlikely to affect its rivals to any great extent. What this means is that each firm in making its decisions does not have to worry how its rivals will react. It assumes that what its rivals choose to do will *not* be influenced by what it does.

  This is known as the assumption of **independence**. (As we shall see later, this is not the case under oligopoly. There we assume that firms believe that their decisions *do* affect their rivals, and that their rivals' decisions will affect them. Under oligopoly we assume that firms are *interdependent*.)
- There is *freedom of entry* of new firms into the industry. If any firm wants to set up in business in this market, it is free to do so.

In these two respects, therefore, monopolistic competition is like perfect competition.

- Unlike perfect competition, however, each firm produces a product or provides a service in some way different from its rivals. As a result it can raise its price without losing all its customers. Thus its demand curve is downward sloping, albeit relatively elastic given the large number of competitors to which customers can turn. This is known as the assumption of **product differentiation.**

TC 10 p 49

Petrol stations, restaurants, hairdressers and builders are all examples of monopolistic competition.

A typical feature of monopolistic competition is that, although there are many firms in the industry, there is only one firm in a particular location. This applies particularly in retailing. There may be several

Definitions

Independence (of firms in a market) Where the decisions of one firm in a market will not have any significant effect on the demand curves of its rivals.

Product differentiation Where one firm's product is sufficiently different from its rivals to allow it to raise the price of the product without customers all switching to the rivals' products. A situation where a firm faces a downward-sloping demand curve.

newsagents in a town, but only one in a particular street. In a sense, therefore, it has a local monopoly. People may be prepared to pay higher prices there to avoid having to go elsewhere.

Pause for thought

An example of monopolistic competition is provided by fast-food restaurants. What other businesses are in competition with fast-food restaurants and what determines the closeness of this competition?

Equilibrium of the firm

Short run

As with other market structures, profits are maximised at the output where $MC = MR$. The diagram is the same as for the monopolist, except that the AR and MR curves are more elastic. This is illustrated in Figure 6.7(a). As with perfect competition, it is possible for the monopolistically competitive firm to make supernormal profit in the short run. This is shown as the shaded area.

TC 2 p 9

Just how much profit the firm will make in the short run depends on the strength of demand: the position and elasticity of the demand curve. The further to the right the demand curve is relative to the average cost curve, and the less elastic the demand curve is, the greater the firm's short-run profit will be. Thus a firm whose product is considerably differentiated from its rivals may be able to earn considerable short-run profits.

TC 10 p 49

Pause for thought

Which of these two items is a petrol station more likely to sell at a discount: (a) oil; (b) sweets? Why?

Long run

If typical firms are earning supernormal profit, new firms will enter the industry in the long run. As new firms enter, they will take some of the customers away from the established firms. The demand for the established firms' product will therefore fall. Their demand (AR) curve will shift to the left, and will continue doing so as long as supernormal profits remain and thus new firms continue entering.

KI 8 p 24

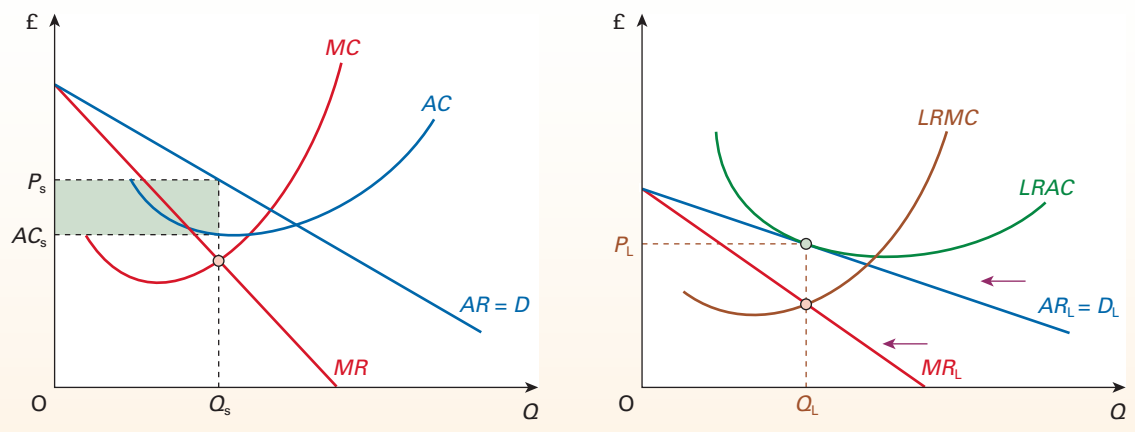
Long-run equilibrium will be reached when only normal profits remain: when there is no further incentive for new firms to enter. This is illustrated in Figure 5.7(b). The firm's demand curve settles at D_L, where it just touches the firm's $LRAC$ curve. Output will be Q_L: where $AR_L = LRAC$. (At any other output, $LRAC$ is greater than AR and thus less than normal profit would be made.)

TC 8 p 37

There is a crucial difference between monopolistic competition and perfect competition in the transition from the short run to the long run, even though in long-run equilibrium, firms in both market structures earn just normal profits.

- Under perfect competition, when new firms enter (or leave) the market, it is the industry supply curve that shifts, which changes the market price and leaves just normal profits.

Figure 6.7 Equilibrium of the firm under monopolistic competition



(a) Short run **(b)** Long run

- Under monopolistic competition, however, the entry of new firms is reflected by shifting an established firm's demand curve inwards and this eliminates the supernormal profits.

As all firms under monopolistic competition are producing a slightly differentiated product, each firm is different and hence we cannot create an industry demand or supply curve. Instead, we have to focus on the effect on a given firm when new firms enter the market.

Non-price competition

One of the biggest problems with the simple model in Figure 6.7 is that it concentrates on price and output decisions. In practice, the profit-maximising firm under monopolistic competition will also need to decide the exact variety of product to produce and how much to spend on advertising it. This will lead the firm to take part in **non-price competition**.

Non-price competition involves two major elements: product development and advertising.

TC 10 p 49

The major aims of *product development* are to produce a product that will sell well (i.e. one in high or potentially high demand) and one that is different from rivals' products (i.e. has a relatively inelastic demand owing to lack of close substitutes). For shops or other firms providing a service, 'product development' takes the form of attempting to provide a service which is better than, or at least different from, that of rivals: personal service, late opening, certain lines stocked, and so on.

The major aim of *advertising* is to sell the product. This can be achieved not only by informing people of the product's existence and availability, but also by trying to persuade them to purchase it. Like product development, successful advertising will both increase demand and also make the firm's demand curve less elastic, since it stresses the specific characteristics of this firm's product over its rivals' (see Box 3.2).

Product development and advertising not only increase a firm's demand and hence revenue; they also involve increased costs. So how much should a firm advertise, say to maximise profits?

For any given price and product, the optimal amount of advertising is where the revenue from *additional* advertising (MR_A) is equal to its cost (MC_A). As long as

TC 2 p 9

$MR_A > MC_A$, additional advertising will add to profit. But extra amounts spent on advertising are likely to lead to smaller and smaller increases in sales. Thus MR_A falls, until $MR_A = MC_A$. At that point no further profit can be made. It is at a maximum.

Pause for thought

Why will additional advertising lead to smaller and smaller increases in sales?

Several problems arise with this analysis. These include:

- The effect of product development and advertising on demand will be difficult for a firm to forecast.
- Product development can affect a firm's future costs. For example, it may impact on production costs or 'aftercare' costs, such as servicing and repair.
- Product development and advertising are likely to have different effects at different prices. Profit maximisation, therefore, will involve the more complex choice of the optimum combination of price, type of product, and level and variety of advertising.

Monopolistic competition and the public interest

Comparison with perfect competition

It is often argued that monopolistic competition leads to a less efficient allocation of resources than perfect competition.

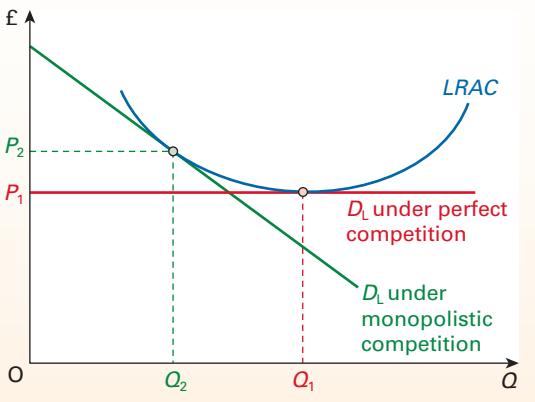
KI 20 p 127

Figure 6.8 compares the long-run equilibrium positions for two firms. One firm is under perfect competition and thus faces a horizontal demand curve. It will produce an output of Q_1 at a price of P_1. The other is under monopolistic competition and thus faces a downward-sloping demand curve. It will produce the lower output of Q_2 at the higher price of P_2. A crucial assumption here is that a firm would have the *same*

Definition

Non-price competition Competition in terms of product promotion (advertising, packaging, etc.) or product development.

Figure 6.8 Long-run equilibrium of the firm under perfect and monopolistic competition



long-run average cost (*LRAC*) curve in both cases. Given this assumption, monopolistic competition has the following disadvantages:

- Less will be sold and at a higher price.
- Firms will not be producing at the least-cost point.

By producing more, firms would move to a lower point on their *LRAC* curve. Thus firms under monopolistic competition are said to have **excess capacity**. In Figure 6.8 this excess capacity is shown as $Q_1 - Q_2$. In other words, monopolistic competition is typified by quite a large number of firms (e.g. petrol stations), all operating at less than optimum output, and thus being forced to charge a price above that which they could charge if they had a bigger turnover.

Pause for thought

Does this imply that if, say, half of the petrol stations were closed down, the consumer would benefit? (Clue: what would happen to the demand curves of the remaining stations?)

So how does this affect the consumer? Although the firm under monopolistic competition may charge a higher price than under perfect competition, the difference may be very small. Although the firm's demand curve is downward sloping, it is still likely to be highly elastic due to the large number of substitutes. Furthermore, the consumer may benefit from monopolistic competition by having a greater variety of products to choose from. Each firm may satisfy some particular requirement of particular consumers.

Comparison with monopoly

The arguments are very similar here to those when comparing perfect competition and monopoly.

On the one hand, freedom of entry for new firms and hence the lack of long-run supernormal profits under monopolistic competition are likely to help keep prices down for the consumer and encourage cost saving. On the other hand, monopolies are likely to achieve greater economies of scale and have more funds for investment and research and development.

Recap

1. Monopolistic competition occurs where there is free entry to the industry and quite a large number of firms operating independently of each other, but where each firm has some market power as a result of producing differentiated products or services.
2. In the short run, firms can make supernormal profits. In the long run, however, freedom of entry will drive profits down to the normal level. The long-run equilibrium of the firm is where the (downward-sloping) demand curve just touches the long-run average cost curve.
3. The long-run equilibrium is one of excess capacity. Given that the demand curve is downward sloping, the point where it just touches the *LRAC* curve will not be at the bottom of the *LRAC* curve. Increased production would thus be possible at lower average cost.
4. Firms under monopolistic competition may engage in non-price competition, in the forms of product development and advertising, in order to maintain an advantage over their rivals.
5. Monopolistically competitive firms, because of excess capacity, may have higher costs, and thus higher prices, than perfectly competitive firms, but consumers may gain from a greater diversity of products.
6. Monopolistically competitive firms may have fewer economies of scale than monopolies and conduct less research and development, but the competition may keep prices lower than under monopoly. Whether there will be more or less choice for the consumer is debatable.

6.5 OLIGOPOLY

What happens if there are just a few firms that dominate the market? Will they compete or get together?

Oligopoly occurs when just a few firms between them share a large proportion of the industry. Some of the best-known companies are oligopolists, including Ford, Nike, Coca-Cola, BP and Apple. Indeed there is a range of industries with powerful oligopolies, including toothbrush manufacturers, supermarkets and energy suppliers.

There are, however, significant differences in the structure of industries under oligopoly and similarly

Definition

Excess capacity (under monopolistic competition) In the long run, firms under monopolistic competition will produce at an output below their minimum-cost point.

significant differences in the behaviour of firms. The firms may produce a virtually identical product (e.g. metals, chemicals, sugar, petrol). Most oligopolists, however, produce differentiated products (e.g. cars, toiletries, soft drinks, electrical appliances). Much of the competition between such oligopolists is in terms of the marketing of their particular brand.

The two key features of oligopoly

Despite the differences between oligopolies, there are two crucial features that distinguish oligopoly from other market structures.

Barriers to entry

Unlike firms under monopolistic competition, there are various barriers to the entry of new firms. These are similar to those under monopoly (see pages 127–31). The size of the barriers, however, will vary from industry to industry. In some cases, entry is relatively easy, whereas in others it is virtually impossible, perhaps due to patent protection or prohibitive research and development costs.

Interdependence of the firms

Because there are only a few firms under oligopoly, each firm has to take account of the others. This means that they are mutually dependent: they are **interdependent**. Each firm is affected by its rivals' actions. If a firm changes the price or specification of its product, for example, or the amount of its advertising, the sales of its rivals will be affected. The rivals may then respond by changing their price, specification or advertising.

KEY IDEA 21

People often think and behave strategically. How you think others will respond to your actions is likely to influence your own behaviour. Firms, for example, when considering a price or product change will often take into account the likely reactions of their rivals.

No firm can therefore afford to ignore the actions and reactions of other firms in the industry.

It is impossible, therefore, to predict the effect on a firm's sales of, say, a change in the price of its product without first making some assumption about the reactions of other firms. Different assumptions will yield different predictions. For this reason there is no single generally accepted theory of oligopoly. Firms may react differently and unpredictably.

Competition and collusion

The interdependence of oligopolists means they are pulled in two different directions.

- Each firm, by carefully studying the market and its rivals' strategy may believe that, by *competing*, it can gain a greater share of industry profits.
- On the other hand, firms may conclude that competition will be destructive and lead to lower profits. So instead, they may prefer to *collude* with each other by making agreements about price, output, product design, etc. By acting together as if they were a monopoly, the firms could take actions that jointly maximise industry profits and share these profits between them.

These two policies are incompatible. The more fiercely firms compete to gain a bigger share of industry profits, the smaller these industry profits will become. For example, price competition drives down the average industry price, while competition through advertising raises industry costs. Either way, industry profits fall.

Sometimes firms will collude; sometimes they will not. The following sections examine first **collusive oligopoly** (both open and tacit), and then **non-collusive oligopoly.**

Collusive oligopoly

When firms under oligopoly engage in collusion, they may agree on prices, market share, advertising expenditure, etc. Such collusion reduces the uncertainty they face. It reduces the fear of engaging in competitive price cutting or retaliatory advertising, both of which could reduce total industry profits and probably each individual firm's profit.

KI 19 p 122

A cartel

A formal collusive agreement is called a **cartel**. The cartel will maximise profits if it acts like a monopoly, with members behaving as if they were a single firm. This is illustrated in Figure 6.9.

Definitions

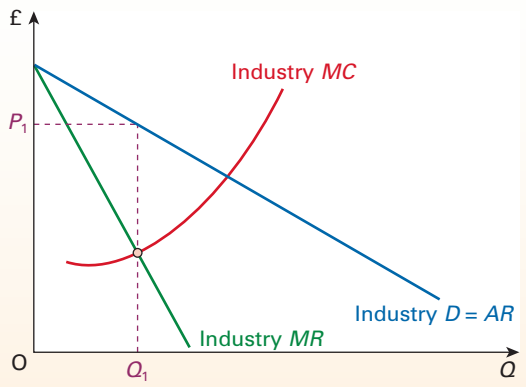
Interdependence (under oligopoly) One of the two key features of oligopoly. Each firm will be affected by its rivals' decisions. Likewise its decisions will affect its rivals. Firms recognise this interdependence. This recognition will affect their decisions.

Collusive oligopoly Where oligopolists agree (formally or informally) to limit competition between themselves. They may set output quotas, fix prices, limit product promotion or development, or agree not to 'poach' each other's markets.

Non-collusive oligopoly Where oligopolists have no agreement between themselves, formal, informal or tacit.

Cartel A formal collusive agreement.

Figure 6.9 Profit-maximising cartel



The total market demand curve is shown with the corresponding market *MR* curve. The cartel's *MC* curve is the *horizontal* sum of the MC curves of its members (since we are adding the *output* of each of the cartel members at each level of marginal cost). Profits are maximised at Q_1 where $MC = MR$. The cartel must therefore set a price of P_1 (at which Q_1 will be demanded).

TC 2 p 9

Having agreed on the cartel price, the members may then compete against each other using *non-price competition,* to gain as big a share of resulting sales (Q_1) as they can.

Alternatively, the cartel members may somehow agree to divide the market between them. Each member would be given a **quota**. The sum of all the quotas must add up to Q_1. If the quotas exceeded Q_1, either there would be output unsold if price remained fixed at P_1, or the price would fall.

But if quotas are to be set by the cartel, how will it decide the level of each individual member's quota? The most likely method is for the cartel to divide the market between the members according to their current market share. That is the solution most likely to be accepted as 'fair'.

Pause for thought

If this 'fair' solution were adopted, what effect would it have on the industry MC curve in Figure 6.9?

In many countries, including the UK, cartels are illegal. They are seen by the government as a means of driving up prices and profits, and thereby as being against the public interest. Where open collusion is illegal, firms may simply break the law, or get round it. Alternatively, firms may stay within the law, but still *tacitly* collude by watching each other's prices and keeping theirs similar. Firms may tacitly 'agree' to avoid price wars or aggressive advertising campaigns.

KI 21 p 140

Tacit collusion

One form of **tacit collusion** is where firms keep to the price that is set by an established leader. The leader may be the largest firm: the firm that dominates the industry. This is known as **dominant firm price leadership**. Alternatively, the price leader may simply be the one that has proved to be the most reliable to follow: the one that is the best barometer of market conditions. This is known as **barometric firm price leadership**.

Dominant firm price leadership. This is a 'sequential game', where one firm (the leader) moves first and then the followers, having observed the leader's choice of price, move second. We will discuss sequential games in more detail in Section 6.6.

But, how does the leader set the price? This depends on the assumptions it makes about its rivals' reactions to its price changes. If it assumes that rivals will simply follow it by making exactly the same percentage price changes up or down, then a simple model can be constructed. This is illustrated in Figure 6.10. The leader assumes that it will maintain a constant market share (say, 50 per cent).

The leader will maximise profits where its marginal revenue is equal to its marginal cost. It knows its current position on its demand curve (say, point *a*) as it knows its current price and how much output it is selling. It then estimates how responsive its demand will be to industry-wide price changes and thus constructs its demand and *MR* curves on that basis. It then chooses to produce Q_L at a price of P_L: at point *l* on its demand curve (where $MC = MR$). Other firms then

Definitions

Quota (set by a cartel) The output that a given member of a cartel is allowed to produce (production quota) or sell (sales quota).

Tacit collusion Where oligopolists take care not to engage in price cutting, excessive advertising or other forms of competition. There may be unwritten 'rules' of collusive behaviour, such as price leadership.

Dominant firm price leadership Where firms (the followers) choose the same price as that set by a dominant firm in the industry (the leader).

Barometric firm price leadership Where the price leader is the one whose prices are believed to reflect market conditions in the most satisfactory way.

BOX 6.3 **OPEC**

The history of the world's most famous cartel

OPEC is probably the best known of all cartels. Set up in 1960 by the five major oil-exporting countries (Saudi Arabia, Iran, Iraq, Kuwait and Venezuela), it currently has 13 members, including Nigeria, Angola, Libya and the United Arab Emirates.

OPEC's main objective is to co-ordinate the supply of oil by its members so as to support an oil price that gives a steady income to producers and a good return on capital. It also aims to ensure a regular supply to consumers.

The years leading up to 1960 saw the oil-producing countries increasingly in conflict with the international oil companies, which extracted oil under 'concessionary agreement'. Under this scheme, oil companies were given the right to extract oil in return for royalties. This meant that the oil-producing countries had little say over output and price levels.

The early years

Despite the formation of OPEC in 1960, it was not until 1973 that control of oil production was effectively transferred from the oil companies to the oil countries, with OPEC deciding how much oil to produce and thereby determining its oil revenue.

OPEC's pricing policy over the 1970s consisted of setting a market price for Saudi Arabian crude (the market leader), and leaving other OPEC members to set their prices in line with this: a form of dominant 'firm' price leadership (see page 207). As demand for oil was both buoyant and inelastic, OPEC could raise prices, which allowed revenue growth.

In 1973/4, after the Arab–Israeli war, OPEC raised the price of oil from around $3 per barrel to over $12. The price was kept at roughly this level until 1979, but with the Iranian Revolution then causing oil output to fall, prices rose from around $15 to $40 per barrel. This, together with the recession of the early 1980s, which can partly be attributed to governments' response to the rising oil prices, caused demand to fall.

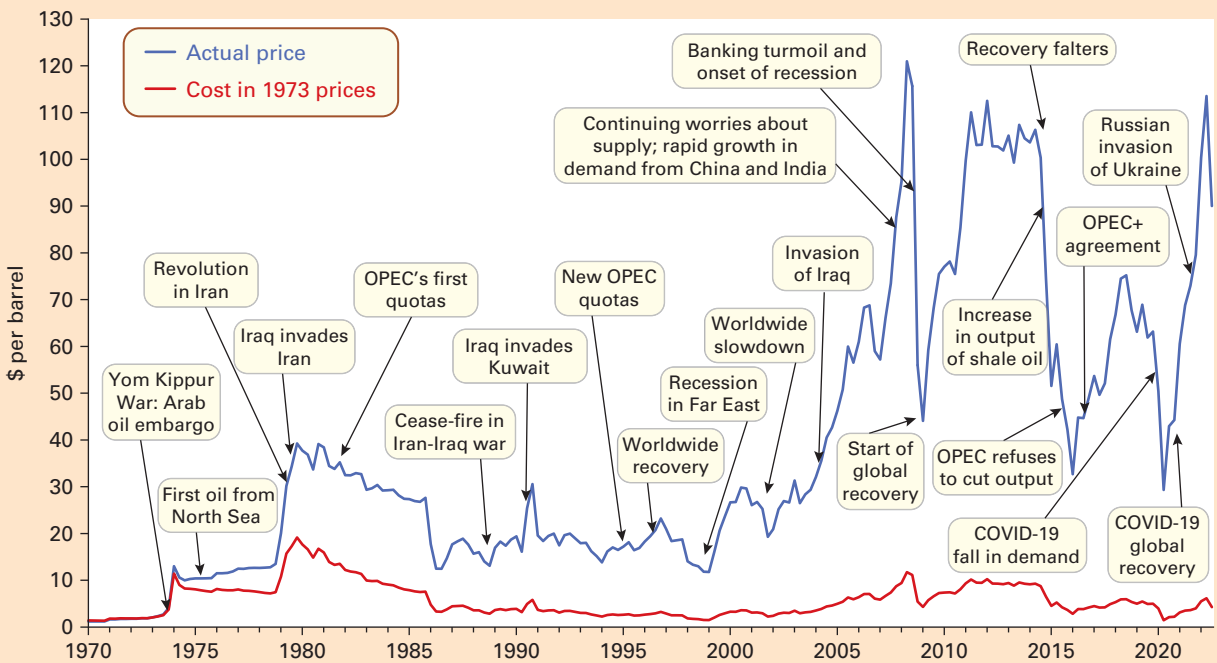
Quotas

Faced by declining demand, OPEC after 1982 agreed to limit output and allocate production quotas in an attempt to keep the price up. However, the cartel was beginning to break down, with the recession reducing demand for oil, growing output from non-OPEC members and some OPEC members exceeding their quotas. With a glut of oil, OPEC could no longer maintain the price.

The world economy boomed in the late 1980s, pushing up the demand for oil. Then in 1990 Iraq invaded Kuwait and the first Gulf War ensued. Supplies from Kuwait and Iraq were cut off, reducing the supply of oil and there was a sharp rise in its price. But with the ending of the war and the early 1990s recession, the price rapidly fell again and only recovered slowly as the world economy started expanding once more.

On the demand side, the development of energy-saving technology plus increases in fuel taxes led to a relatively slow growth in consumption. On the supply side, the growing proportion of output supplied by non-OPEC members, plus the

Oil prices (Brent crude, quarterly average)



Sources: Nominal oil price data from *World Commodity Price Data (The Pink Sheet)*, *Commodity Markets* (World Bank); Price Index from *Data Extracts* (OECD)

adoption in 1994 of a relatively high OPEC production ceiling of 24½ million barrels per day, meant that supply more than kept pace with demand.

In the late 1990s, the recession in the Far East reduced demand by some 2 million barrels per day, causing prices to fall once more. OPEC's members intervened by cutting production by 4.3 million barrels per day in an attempt to push the price back up to around $18–20 per barrel.

In late 2001, five of the major oil producers outside of the cartel, including Russia, formed an alliance with OPEC members, whereby they all agreed to cut output, aiming to push up and stabilise prices at around $25 per barrel. This type of alliance was the first of its kind in the oil industry and it gave control of oil back to OPEC. However, the Iraq War in 2003 created supply problems and although OPEC tried to relax its quotas, rising demand from India and China meant that OPEC could not adjust supply sufficiently quickly to make any real difference to the price.

The 2008 financial crisis and a new competitor

The financial crisis pushed oil prices down dramatically: falling from a high of $147 per barrel in July 2008 to a low of $34 in December 2008. The lower prices, while good for consumers, were potentially damaging for investment in both oil exploration/development and alternative energy supplies.

OPEC responded to the falling price by announcing cuts in production, totalling some 14 per cent between August 2008 and January 2009. But with OPEC now producing less than a third of global oil output, this represented less than 5 per cent of global production and consequently had little effect on the price.

Prices recovered and remained stable between mid-2011 and mid-2014 at around $100 to $120 per barrel. But then OPEC faced a new competitor in the form of US shale oil production. With a new source of supply, oil prices fell – from $112 per barrel in June 2014 to just $30 per barrel in February 2016.

OPEC responded, not by cutting price, but by announcing that it would retain output at existing levels. What it was relying on was the fact that production from shale oil wells, although often involving low marginal costs, often lasts only two or three years. Investment in new shale oil wells, by contrast, tends to be relatively expensive. By OPEC maintaining production, it was hoping to use its remaining market power to reduce supply of competitors over the medium to long term.

However, with revenues from oil falling so dramatically, OPEC plus 10 non-OPEC producers, including Russia, Mexico and Azerbaijan (known as the OPEC+ Alliance), reached an agreement in December 2016 to cut production. The oil price immediately started to increase again. However, the higher oil prices made US shale production profitable once again. A number of rigs that had temporarily shut down became operational again and production increased to over 9.1 million barrels a day in early 2020, around 9 per cent of total global output. The OPEC+ Alliance, responded by extending its quota agreements and agreeing to future production cuts.

COVID-19 and Russia's invasion of Ukraine

As countries entered lockdowns in 2020, global demand fell by over 30 per cent. Although Saudi Arabia wanted the OPEC+ Alliance to cut production significantly to keep prices up, Russia opposed the cuts because it would help keep US shale producers in business, and so no agreement was reached.

A price war began with production increasing and, together with the continued fall in demand due to COVID-19, there was a dramatic fall in prices from over $50 per barrel on 5 March to under $32 per barrel on 9 March 2020. Eventually the OPEC+ Alliance agreed to production cuts, but the downward trend in prices continued until lockdowns eased.

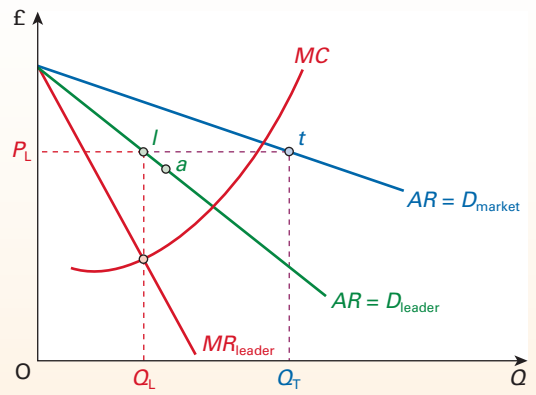
Throughout 2021 and in early 2022 prices rebounded, despite the development of new variants of COVID, as demand continued to rise and the global economy began to recover. This contributed to record petrol and diesel prices and inflationary pressures, which caused significant problems for many households.

On 24 February 2022, Russia invaded Ukraine. Severe sanctions were imposed on Russia by the USA, the EU and many other countries. With Russia being the world's second largest oil producer (behind the USA and just ahead of Saudi Arabia) and a key oil exporter, it was unsurprising that oil prices rose rapidly and passed $100 per barrel – the highest level in over seven years. It may be that already high oil prices are going to continue to rise, but much of this will depend on the duration of the conflict and the speed that other oil producers increase production to replace Russian oil. As the chart shows, however, oil prices began to fall from mid-2022 as the world economy slowed down and, with it, the demand for oil.

1. *What conditions facilitate the formation of a cartel? Which of these conditions were to be found in the oil market in (a) the early 1970s; (b) the mid-1980s; (c) the mid 2000s; (d) the mid-to-late 2010s?*
2. *Could OPEC have done anything to prevent the long-term decline in real oil prices seen from 1981 to 2002?*
3. *Do the increased demand from China and India and the sanctions against Russia imply that the era of cheap energy is over? What impact could technology have in the long run on (a) demand; (b) supply?*

Download monthly price data on commodity markets from the World Bank. Create a chart showing the annual rate of oil price inflation from the early 1970s. Write a short commentary summarising the patterns observed in oil price inflation.

Figure 6.10 A price leader aiming to maximise profits for a given market share



follow that price. Total market demand is Q_T, with followers supplying that portion of the market not supplied by the leader: namely, $Q_T - Q_L$.

There is one problem with this model: that is the assumption that the followers will want to maintain a constant market share. It is possible that if the leader raises its price, the followers may want to supply more at this new price. On the other hand, the followers may decide merely to maintain their market share for fear of invoking retaliation from the leader, in the form of price cuts or an aggressive advertising campaign.

Barometric firm price leadership. A similar exercise can be conducted by a barometric firm. Although the firm is not dominating the industry, its price will be followed by the others. It merely tries to estimate its demand and *MR* curves – assuming, again, a constant market share – and then produces where $MR = MC$ and sets price accordingly.

In practice, which firm is taken as the barometer may frequently change. Whether we are talking about oil companies, car producers or banks, any firm may take the initiative in raising prices. If the other firms are merely waiting for someone to take the lead – say, because costs have risen – they will all quickly follow suit. For example, if one of the bigger building societies or banks raises its mortgage rates by 1 per cent, this is likely to stimulate the others to follow suit.

Other forms of tacit collusion. An alternative to having an established leader is for there to be an established set of simple 'rules of thumb' that everyone follows.

One such example is *average cost pricing* (see also Section 5.5). Instead of equating *MC* and *MR*, producers simply add a certain percentage for profit on top of average costs. Thus, if average costs rise by 10 per cent, prices will automatically be raised by 10 per cent. This is a particularly useful rule of thumb in times of inflation, when all firms will be experiencing similar cost increases.

Pause for thought

If a firm has a typically shaped average cost curve and sets prices 10 per cent above average cost, what will its supply curve look like?

Another rule of thumb is to have certain **price benchmarks**. Thus clothes may sell for £24.99, £49.95 or £99.95 (but not, say, for £23.31 or £102.42). If costs rise, then firms simply raise their price to the next benchmark, knowing that other firms will do the same.

Rules of thumb can also be applied to advertising (e.g. you do not criticise other firms' products, only praise your own); or to the design of the product (e.g. lighting manufacturers tacitly agreeing not to bring out an everlasting light bulb).

Factors favouring collusion

Collusion between firms, whether formal or tacit, is more likely when firms can clearly identify with each other or some leader and when they trust each other not to break agreements. It is easier for firms to collude if the following conditions apply:

- There are only very few firms, all well known to each other.
- They are open with each other about costs and production methods.
- They have similar production methods and average costs, and are thus likely to want to change prices at the same time and by the same percentage.
- They produce similar products and can thus more easily reach agreements on price.
- There is a dominant firm.
- There are significant barriers to entry and thus there is little fear of disruption by new firms.
- The market is stable. If industry demand or production costs fluctuate wildly, it will be difficult to make agreements, partly due to difficulties in predicting and partly because agreements may frequently have to be amended. There is a particular problem in a declining

Definition

Price benchmark A price that is typically used. Firms, when raising a price, will usually raise it from one benchmark to another.

market where firms may be tempted to undercut each other's price in order to maintain their sales.
- There are no government measures to curb collusion.

Pause for thought

In which of the following industries is collusion likely to occur: bricks, beer, margarine, cement, crisps, washing powder, USB flash drives, carpets?

Non-collusive oligopoly: the breakdown of collusion

In some oligopolies, there may be only a few (if any) factors favouring collusion. In such cases, the likelihood of price competition is greater.

Even if there is collusion, there will always be the temptation for individual oligopolists to 'cheat', by cutting prices or by selling more than their allotted quota. The danger, of course, is that this would invite retaliation from the other members of the cartel, with a resulting price war. Price would then fall and the cartel could well break up in disarray.

When considering whether to break a collusive agreement, even if only a tacit one, a firm will ask: (1) 'How much can we get away with without inviting retaliation?' and (2) 'If a price war does result, will we be the winners? Will we succeed in driving some or all of our rivals out of business and yet survive ourselves, and thereby gain greater market power?'

The position of rival firms, therefore, is rather like that of generals of opposing armies or the players in a game. It is a question of choosing the appropriate *strategy*: the strategy that will best succeed in outwitting your opponents. The strategy a firm adopts will, of course, be concerned not just with price but also with advertising and product development.

KI 21 p 140

Non-collusive oligopoly: assumptions about rivals' behaviour

Even though oligopolists might not collude, they will still need to take account of rivals' likely behaviour when deciding their own strategy. In doing so they will probably look at rivals' past behaviour and make assumptions based on it. There are three well-known models, each based on a different set of assumptions.

Assumption that rivals produce a given quantity: the Cournot model

One assumption is that rivals will produce a particular *quantity*. This is most likely when the market is stable and the rivals have been producing a relatively constant quantity for some time. The task, then, for the individual oligopolist is to decide its own price and quantity given the presumed output of its competitors.

The earliest model based on this assumption was developed by the French economist Augustin Cournot in 1838. The **Cournot model** (which is developed in Web Appendix 6.2) takes the simple case of just two firms (a **duopoly**) producing an identical product: for example, two electricity generating companies supplying the whole country.

TC 4 p 13

This is illustrated in Figure 6.11 which shows the profit-maximising price and output for Firm A. The total market demand curve is shown as D_M. Assume that Firm A believes that its rival, Firm B, will produce Q_{B1} units. Thus Firm A perceives its own demand curve (D_{A1}) to be Q_{B1} units less than total market demand. In other words, the horizontal gap between D_M and D_{A1} is Q_{B1} units. Given its perceived demand curve of D_{A1}, its marginal revenue curve will be MR_{A1} and the profit-maximising output will be Q_{A1}, where $MR_{A1} = MC_A$. The profit-maximising price will be P_{A1}.

If Firm A believed that Firm B would produce *more* than Q_{B1}, its perceived demand and *MR* curves would be further to the left and the profit-maximising quantity and price would both be lower.

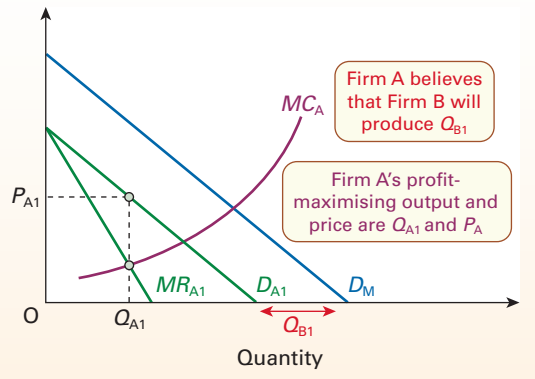
At the same time as Firm A makes an assumption about Firm B's output, Firm B will also be making an assumption about how much it thinks Firm A will produce. This is therefore a 'simultaneous game',

Definitions

Cournot model A model of duopoly where each firm makes its price and output decisions on the assumption that its rival will produce a particular quantity.

Duopoly An oligopoly where there are just two firms in the market.

Figure 6.11 The Cournot model of duopoly: Firm A's profit-maximising position



as both firms are making their decisions at the same time and, crucially, cannot observe the action of the other firm. We discuss this in Section 6.6.

Profits in the Cournot model. Industry profits will be *less* than under a monopoly or a cartel. The reason is that price will be lower than the monopoly price. This can be seen from Figure 6.11. If this were a monopoly, then to find the profit-maximising output, we would need to construct an *MR* curve corresponding to the market demand curve (D_M). This would intersect with the *MC* curve at a higher output than Q_{A1} and a *higher* price (given by D_M).

Nevertheless, profits in the Cournot model will be higher than under perfect competition, since price is still above marginal cost.

Assumption that rivals set a particular price: the Bertrand model

An alternative assumption is that rival firms set a particular price and stick to it. This scenario is more realistic when firms do not want to upset customers by frequent price changes or want to produce catalogues which specify prices. The task, then, for a given oligopolist is to choose its own price and quantity in the light of the prices set by rivals.

The most famous model based on this assumption was developed by another French economist, Joseph Bertrand in 1883. Bertrand again took the simple case of a duopoly, but its conclusions apply equally to oligopolies with three or more firms.

The outcome is one of price cutting until all supernormal profits are competed away. The reason is simple. If Firm A assumes that its rival, Firm B, will hold price constant, then Firm A should undercut this price by a small amount and as a result gain a large share of the market. At this point, Firm B will be forced to respond by cutting its price. What we end up with is a price war until price is forced down to the level of average cost, with only normal profits remaining.

Definitions

Nash equilibrium The position resulting from everyone making their optimal decision based on their assumptions about their rivals' decisions. Without collusion, there is no incentive for any firm to move from this position.

Takeover bid Where one firm attempts to purchase another by offering to buy the shares of that company from its shareholders.

Kinked demand theory The theory that oligopolists face a demand curve that is kinked at the current price, demand being significantly more elastic above the current price than below. The effect of this is to create a situation of price stability.

As with the Cournot model above, this is also a simultaneous-move game, except here the variable of interest is price. The supermarket industry is a good example of a market where price wars are a constant feature.

Nash equilibrium. The equilibrium outcome in either the Cournot or Bertrand models is not in the *joint* interests of the firms. In each case, total profits are less than under a monopoly or cartel. But, in the absence of collusion, the outcome is the result of each firm doing the best it can, given its assumptions about what its rivals are doing. The resulting equilibrium is known as a **Nash equilibrium** after John Nash, a US mathematician (and subject of the film *A Beautiful Mind*) who introduced the concept in 1951. We will return to this concept in Section 6.6.

In practice, when competition is intense, as in the Bertrand model, the firms may seek to collude long before profits have been reduced to a normal level. Alternatively firms may put in a **takeover bid** for their rival(s).

The kinked demand-curve assumption

In 1939 a theory of non-collusive oligopoly was developed simultaneously on both sides of the Atlantic: in the USA by Paul Sweezy and in Britain by R.L. Hall and C.J. Hitch. This **kinked demand theory** has since become perhaps the most famous of all theories of oligopoly. The model seeks to explain how it is that, even when there is no collusion at all between oligopolists, prices can nevertheless remain stable.

The theory is based on two asymmetrical assumptions.

TC 5 p 17

- If an oligopolist cuts its price, its rivals will feel forced to follow suit and cut theirs, to prevent losing customers to the first firm.
- If an oligopolist raises its price, however, its rivals will *not* follow suit since, by keeping their prices the same, they will thereby gain customers from the first firm.

On these assumptions, each oligopolist will face a demand curve that is *kinked* at the current price and output (see Figure 6.12). A rise in price will lead to a large fall in sales as customers switch to the now relatively lower-priced rivals. The firm will thus be reluctant to raise its price. Demand is relatively elastic above the kink. On the other hand, a fall in price will bring only a modest increase in sales, since rivals lower their prices too and therefore customers do not switch. The firm will thus also be reluctant to lower its price. Demand is relatively inelastic below the kink. Thus oligopolists will be reluctant to change prices at all.

TC 10 p 49

This price stability can be shown formally by drawing in the firm's marginal revenue curve, as in Figure 6.13.

Figure 6.12 Kinked demand curve for a firm under oligopoly

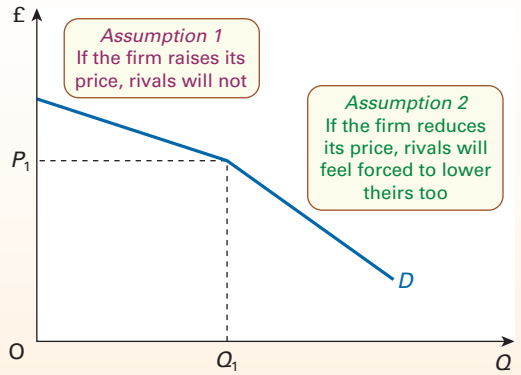
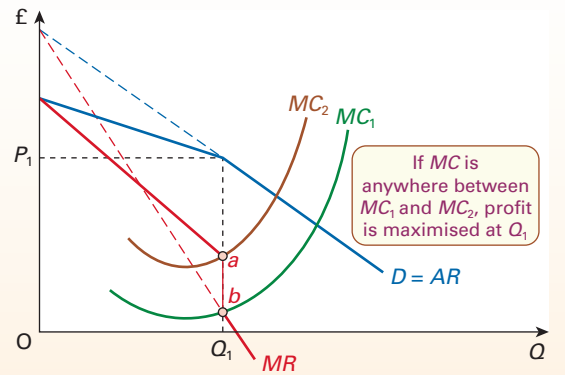


Figure 6.13 Stable price under conditions of a kinked demand curve



To see how this is done, imagine dividing the diagram into two parts either side of Q_1. At quantities less than Q_1 (the left-hand part of the diagram), the *MR* curve will correspond to the shallow part of the *AR* curve. At quantities greater than Q_1 (the right-hand part), the *MR* curve will correspond to the steep part of the *AR* curve. To see how this part of the *MR* curve is constructed, imagine extending the steep part of the *AR* curve back to the vertical axis. This and the corresponding *MR* curve are shown by the dotted lines in Figure 6.13.

As you can see, there will be a gap between points *a* and *b*. In other words, there is a vertical section of the *MR* curve between these two points. Profits are maximised where $MC = MR$. Thus, if the *MC* curve lies anywhere between MC_1 and MC_2 (i.e. between points *a* and *b*), the profit-maximising price and output will be P_1 and Q_1. Thus prices will remain stable *even with a considerable change in costs*.

Oligopoly and the consumer

If oligopolists act collusively and jointly maximise industry profits, they will in effect be acting together as a monopoly. In such cases, prices may be very high. This is clearly not in the best interests of consumers.

KI 19 p 122

Furthermore, in two respects, oligopoly may be more disadvantageous than monopoly.

- Depending on the size of the individual oligopolists, there may be less scope for economies of scale to mitigate the effects of market power.
- Oligopolists are likely to engage in much more extensive advertising than a monopolist. This will raise costs. Consumers could thus end up paying higher prices, though it may lead to product development and better information about the product's characteristics.

These problems will be less severe, however, if oligopolists do not collude, if there is some degree of price competition and if barriers to entry are weak. For example, in the Bertrand model, prices end up being set at the perfectly competitive level.

Moreover, the power of oligopolists in certain markets may to some extent be offset if they sell their product to other powerful firms. Thus oligopolistic producers of baked beans or soap powder sell a large proportion of their output to giant supermarket chains, which can use their market power to keep down the price at which they purchase these products. This phenomenon is known as **countervailing power**.

Pause for thought

Which of the following are examples of effective countervailing power?
(a) Tour operators purchasing seats on charter flights; (b) A large office hiring a photocopier from Xerox; (c) Marks & Spencer buying clothes from a garment manufacturer; and (d) A small village store (but the only one for miles around) buying food from a wholesaler.

In some respects, oligopoly has *advantages* to society over other market structures.

Definition

Countervailing power Where the power of a monopolistic/oligopolistic seller is offset by powerful buyers which can prevent the price from being pushed up.

- Oligopolists, like monopolists, can use part of their supernormal profit for research and development. Unlike monopolists, however, oligopolists will have a considerable *incentive* to do so. If the product design is improved, this may allow the firm to capture a larger share of the market, and it may be some time before rivals can respond with a similarly improved product. If, in addition, costs are reduced by technological improvement, the resulting higher profits will improve the firm's capacity to withstand a price war.

TC 6 p 24

- Non-price competition through product differentiation may result in greater choice for the consumer. Take the case of tablets or mobile phones. Non-price competition has led to a huge range of different products of many different specifications, each meeting the specific requirements of different consumers.

It is difficult to draw any general conclusions, since firms within oligopolies differ so much in their performance.

BOX 6.4 **BUYING POWER** CASE STUDIES & APPLICATIONS

What's being served up by the UK grocery sector?

Over the past few years there has been increasing concern about the power of large supermarket chains in the UK. This has resulted in a number of investigations by the country's competition authorities. Many of these investigations have focused on anti-competitive practices, such as rival chains agreeing not to set up in the same town and price collusion on some staple products.

However, more recently, focus has turned to the supermarkets' power not as sellers, but as buyers. If a wholesale manufacturer of ready-meals, or a supplier of sausages, wants to reach a wide customer base, it will need to deal with one or more of the eight largest supermarket chains, which control some 93 per cent of the grocery market. A market like this, where there are a few large purchasers of goods and services, is known as an ***oligopsony***. (A single large buyer of goods, services or factors of production is known as a *monopsony* and we look at this in Chapter 7, pages 168–9.)

Buying power

Over the years, a number of unfair practices by the supermarkets towards their suppliers have been identified. These include retrospectively changing contracts, delaying payments, forcing suppliers to fund special offers such as 'buy one, get one free' and asking for very substantial payments in order to be included on 'preferred supplier' lists.

Following a lengthy investigation, the Competition Commission (a predecessor to the Competition and Markets Authority) found that the supermarkets were passing on excessive risks and unexpected costs to their suppliers. As a consequence, a stronger Grocery Supplies Code of Practice (GSCP) was introduced in 2009. This recognised the power that large grocery retailers wield over their smaller suppliers and outlawed the practices detailed above.

A Groceries Code Adjudicator

In 2013, the government appointed a 'Groceries Code Adjudicator' (GCA), Christine Tacon, to ensure that supermarkets were complying with the GSCP. In 2016, the GCA concluded that Tesco had not complied with the code and had 'knowingly delayed paying money to suppliers . . . to improve its own financial position'. Morrisons also broke the code by requiring suppliers to make lump-sum payments even though the supply contracts didn't require this.

Tacon's approach was about collaboration and nudging the supermarkets to get in line, albeit with penalties imposed where necessary. Some concessions were made during COVID-19 to help supermarkets provide supplies to households, but Tacon was forced to intervene to stop supermarkets from having aspects of the GSCP waived. In April 2020 she noted that the code requires suppliers to be given 'reasonable notice' regarding de-listing and variation of supply agreements but that the code had sufficient flexibility to determine what is 'reasonable' on a case-by-case basis, and that it could also take the effects of the pandemic into account.

In 2020, a new Adjudicator was appointed. Suppliers may hope he takes a stronger approach, while supermarkets may fear a move away from collaboration and self-regulation.

Conclusions

Despite the supermarkets' power both as buyers and sellers, it is too simplistic to assume that the industry outcomes are all bad. Suppliers have been exploited, but the degree of competition between the largest supermarkets has kept food prices down and new products and offers have provided more convenience for many people, particularly those working full-time.

The grocery market is therefore a good example of a sector where growth and market power can be identified as both beneficial and harmful to other economic agents.

1. *Explain why manufacturers of food products continue to supply supermarkets, despite concerns that they are not always treated fairly.*
2. *Is the supermarket sector an oligopoly or monopolistically competitive? Justify your answer.*

Visit the website of the Competition and Markets Authority. Under CMA cases, search for cases/investigations involving the grocery sector. Choose a particular case and summarise the economic ideas and principles relevant to the case.

Definition

Oligopsony A market with just a few buyers (or employers in the case of labour markets).

Recap

1. An oligopoly is where there are just a few firms in the industry with barriers to the entry of new firms. Firms recognise their mutual dependence.
2. Oligopolists want to maximise their joint profits. This tends to make them collude to keep prices high. On the other hand, they want the biggest share of industry profits for themselves. This tends to make them compete.
3. They are more likely to collude if there are few of them; if they are open with each other; if they have similar products and cost structures; if there is a dominant firm; if there are significant entry barriers; if the market is stable; and if there is no government legislation to prevent collusion.
4. Collusion can be open or tacit.
5. A formal collusive agreement is called a 'cartel'. A cartel aims to act as a monopoly. It can set price and leave the members to compete for market share, or it can assign quotas. There is always a temptation for cartel members to 'cheat' by undercutting the cartel price if they think they can get away with it and not trigger a price war.
6. Tacit collusion can take the form of price leadership. This is where firms follow the price set by either a dominant firm in the industry or one seen as a reliable 'barometer' of market conditions. Alternatively, tacit collusion can simply involve following various rules of thumb such as average cost pricing and benchmark pricing.
7. Even when firms do not collude, they will still have to take into account their rivals' behaviour.
8. In the Cournot model firms assume that their rivals' output is given and then choose the profit-maximising price and output in the light of this assumption. The resulting price and profit are lower than under monopoly, but still higher than under perfect competition.
9. In the Bertrand model firms assume that their rivals' price is given. This will result in prices being competed down until only normal profits remain.
10. In the kinked demand-curve model, firms are likely to keep their prices stable unless there is a large shift in costs or demand.
11. Whether consumers benefit from oligopoly depends on the particular oligopoly and how competitive it is; whether there is any countervailing power; whether the firms engage in extensive advertising and of what type; whether product differentiation results in a wide range of choice for the consumer; and how much of the profits are ploughed back into research and development.

6.6 GAME THEORY

Thinking strategically about rivals

The interdependence between oligopolists requires firms to think strategically, making assumptions about rivals' behaviour before taking decisions. **Game theory** is used by economists to examine the best strategy that a firm can adopt and how this may affect market outcomes.

KI 21 p 140

This section will show how game theory can provide useful insights into firms' behaviour in oligopolistic markets. It can also be applied to a broad range of other situations, including the negotiations that took place over the UK's exit from the European Union, international attempts to reach agreement over reducing CO_2 emissions and the war between Russia and the Ukraine.

Pause for thought

Can you think of other examples within economics where game theory could help to analyse the choices made by individuals, groups or organisations?

Simultaneous single-move games

A firm's profit-maximising strategy in a competitive oligopoly market depends, in part, on how it thinks its rivals will react to its decisions on prices, output, product development, advertising, etc. If this competition is a one-off event (such as firms competing for a specific contract) then it can be modelled as a **simultaneous single-move game**. This type of game is also called a single-period or one-shot game.

Definitions

Game theory (or the theory of games) The study of alternative strategies that oligopolists may choose to adopt, depending on their assumptions about their rivals' behaviour.

Simultaneous single-move game A game where each player has just one move, where each player plays at the same time and acts without knowledge of the actions chosen by other players.

Table 6.2 Profits for firms X and Y at different prices

		X's price £2.00	X's price £1.80
Y's price	£2.00	A £10m, £10m	B £5m, £12m
Y's price	£1.80	C £12m, £5m	D £8m, £8m

Simple dominant strategy games

Consider the case where there are just two firms with identical costs, products and demand. They are both considering which of two alternative prices to charge. Table 6.2 shows typical profits they could each make.

Let us assume that at present both firms (X and Y) are charging a price of £2 and that they are each making a profit of £10 million, giving a total industry profit of £20 million. This is shown in the top left-hand cell (A).

KI 21 p 140

Now assume they are both (independently) considering reducing their price to £1.80. In making this decision they will need to take into account what their rival might do, and how this will affect them. Let us consider X's position. In our simple example there are just two things that its rival, firm Y, might do. Either Y could cut its price to £1.80, or it could leave its price at £2. What should X do?

To answer this question we need to take each of firm Y's two possible actions and look at firm X's best response to each. If we assume that firm Y chooses a price of £2, firm X could decide to keep its price at £2 giving it £10m in profit. This is shown by cell A. Alternatively, firm X could cut its price to £1.80 and earn £12m in profit, in cell B. Firm X's best response is therefore to cut price to £1.80, preferring a profit of £12m to one of £10m.

What happens if we now assume that firm Y charges £1.80 – how should firm X best respond? If firm X charged £2, we would end up in cell C and firm X would earn only £5m in profit. On the other hand, firm X could also cut its price to £1.80, moving us to cell D and it would earn £8m profit. By comparing these two profit outcomes, we can see that firm X's best response to firm Y lowering its price to £1.80 is to cut its own price to £1.80 as well, preferring a profit of £8m to a profit of £5m.

Note that firm Y will argue along similar lines, cutting price to £1.80 as well, no matter what it assumes that firm X will do.

This game is known as a **dominant strategy game** since, no matter what the other firm decides, each firm will make the same decision to cut price. As we saw in the Cournot and Bertrand models above, the equilibrium outcome where there is no collusion between the firms (cell D in this game) is known as a *Nash equilibrium*.

TC 11 p 61

However, it is important to note that the profits earned by each firm in the Nash equilibrium (cell D) are lower than they would have been had the firms colluded and charged the higher price (cell A). Each firm would have earned £10 million (cell A). But if they did collude, both would be tempted to cheat and cut prices. This is known as the **prisoners' dilemma**. An example is given in Box 6.5.

Pause for thought

Is the incentive structure in the single-move prisoners' dilemma game in the interests of society?

More complex simultaneous single-move games

More complex 'games' can be devised with more than two firms, many alternative prices, differentiated products and various forms of non-price competition (e.g. advertising). We may also see 'games', where the best response for each firm depends on the assumptions made, meaning there is no dominant strategy. Consider the pay-off matrix in Table 6.3. This is similar to the example in Table 6.2, but with a different profit structure.

Consider firm Y's position. If it assumes that firm X will choose a price of £2, its best response is to charge £1.80, earning £20 million in profits as shown in cell C. However, if it assumes that firm X will choose a price of £1.80, its best response is to charge £2, earning £15 million in profits as shown in cell B. Hence, its best response changes depending on what price it thinks firm X will charge.

Accurately predicting firm X's decision is important for firm Y if it wants to maximise its profits. If its belief turns out to be wrong, it will make less profit. What is the most effective way of anticipating what your rival will do? The answer is for firm Y to try to examine the decision from the perspective of firm X. If it does, it will see that firm X has a dominant strategy.

If firm Y charges £2 it can see that firm X's best response is to charge £1.80 (making a profit of £6m

Definitions

Dominant strategy game Where the *same* policy is suggested by different strategies.

Prisoners' dilemma Where two or more firms (or people), by attempting independently to choose the best strategy for whatever the other(s) are likely to do, end up in a worse position than if they had co-operated in the first place.

Table 6.3 A more complicated game

		X's price: £2.00	X's price: £1.80
Y's price: £2.00		A £18m, £5m	B £15m, £6m
Y's price: £1.80		C £20m, £3m	D £12m, £4m

rather than £5m). If firm Y charges £1.80 firm X's best response is to also charge £1.80 (making a profit of £4m rather than £3m). Therefore, firm Y can predict with a high level of certainty that firm X will charge £1.80 – its dominant strategy. Firm Y's best response, therefore, is to charge £2.00 and make a profit of £15m rather than £12m. This combination of prices in cell B is the equilibrium in the game.

In many situations, firms will have a number of different options open to them and a number of possible reactions by rivals. Such games can become highly complex.

The better the firm's information about (a) its rivals' costs and demand, (b) the likely reactions of rivals to its actions and (c) the effects of these reactions on its own profit, the better the firm's 'move in the game' is likely to be. It is similar to a card game: the more you know about your opponents' cards and how your opponents are likely to react to your moves, and the better you can calculate the effects of their moves on you, the better your moves in the game are likely to be.

Repeated simultaneous-move games

The previous analysis of simultaneous single-move games gives some useful insights but instances of one-off interactions are relatively unusual. In most real-world settings, firms in oligopolistic markets compete against one another on a repeated basis. Decisions about pricing, advertising, product development, etc. are made continually over the months and years that firms are in business. For example, Apple and Samsung launch new versions of their smartphone handsets on an annual basis.

Do the predicted outcomes of single-move games remain the same when the game is repeated?

The big difference between a single-move game and a repeated game is that each firm can now see what its rivals did in previous periods. This creates the possibility that whatever firms choose to do in one period might have an impact on the behaviour of their rivals, and hence their own profits, in later periods. Hence, decisions that generate higher profits today could lead to lower profits in the future. The potential impact of this trade-off can be illustrated by using the prisoners' dilemma example in Table 6.2.

Assume that firms X and Y have an agreement to charge £2 (or simply follow each other's lead in doing so). Each firm will thus continue charging £2 as long as the other firm does too. If, however, firm Y ever cut its price to £1.80 in one period, then firm X would do the same in all future periods. Once this has happened, no matter what firm Y then does, firm X is unlikely to charge a price of £2 ever again. In game theory, this new strategy employed by firm Y is known as the **trigger strategy**.

The profit profile for firm Y of following two different pricing strategies is illustrated in Figure 6.14. By following the same dominant strategy as in the one-move game and charging a price of £1.80 (i.e. breaking the agreement), firm Y can increase its profit in the first period from £10 million to £12 million. The downside of this strategy is that its profits in all future periods will fall to £8 million as firm X responds by also charging £1.80. This is illustrated by the profit profile of a → b → c → d in Figure 6.14.

Alternatively, firm Y could stick to the agreement and charge £2.00 in the first period. Its profit of £10 million is £2 million lower than it would have earned by charging £1.80. However, as long as it maintains its price at this level, firm X will also charge £2.00. Firm Y's profits in all future periods will thus be £10 million as opposed to £8 million. This is shown by the profit profile e → f → g.

After a while, both Y and X will realise that the Nash equilibrium (£1.80) is not to the advantage of either. This may persuade them to set up a stronger collusive agreement to restore prices to £2. This is more likely if the firms value future profits quite highly, they compete against each other frequently and they can readily observe that the rival is charging the lower price.

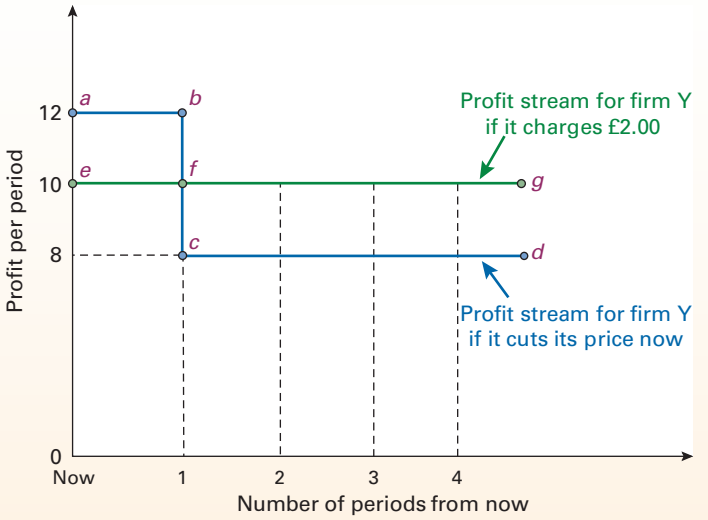
Tit-for-tat strategy. In many situations, it is not necessary for there to be an agreement for a price war to be avoided. If a firm fears that cutting its price will trigger price cutting by rivals, it will be reluctant to cut prices first. On the other hand, it will want to make it clear to rivals that, should they cut prices, it will respond by doing so too. This is known as *a* **tit-for-tat strategy**: 'If you cut price, we will too.' The hope is that the rival will realise this and will not cut prices.

Definitions

Trigger strategy Once a firm observes that its rival has broken some agreed behaviour it will never co-operate with them ever again.

Tit-for-tat strategy Where a firm will always respond to a price cut by a rival by doing the same.

Figure 6.14 Profits for firms Y in a repeated game



Backwards induction and movement to the Nash equilibrium. Another factor increasing the likelihood of firms co-operating and charging higher prices is uncertainty over how long current designs and costs will last. If this point in time can be identified, then firms will be more inclined to follow their dominant strategies of charging £1.80 throughout.

To understand why, consider the likely outcome in the last period of competition (i.e. the last time prices are set before any changes in product design, costs, etc.). At this point there is no future to affect since firms will never compete against one another again with the same product. Hence, there is no longer any incentive for each firm to charge £2 in order to persuade its rival to do the same in future periods. The last period is then effectively the same as a simultaneous single-move game and both firms are highly likely to follow their dominant strategies of charging £1.80.

If we assume that firms work backwards from the last period to think about outcomes in earlier periods, then they engage in a process known as **backwards induction**. If both firms realise in the last-but-one period that they cannot influence what their rival will do in the last period of competition then their best strategy is also to charge £1.80. If they keep following the same line of reasoning, they will both charge £1.80 in every period.

Therefore, the chances of the firms charging a higher price is much greater when they do not know when the current type of competition between them will come to an end.

Definitions

Backwards induction A process by which firms think through the most likely outcome in the last period of competition and then work backwards step by step thinking through the most likely outcomes in earlier periods of competition.

Sequential-move games One firm (the first mover) makes and implements a decision. Rival firms (second movers) can observe the actions taken by the first mover before making their own decisions.

Sequential-move games

In simultaneous games firms take decisions at the same time without seeing the decision of other firms. However, in many real-world competitive environments, one firm (the first mover) makes and implements a decision (i.e. it produces a certain output, sets a particular price or introduces a new product) *before* its rivals (the second movers). The second movers are then able to observe the actions of the first mover before deciding on their best response. These strategic environments can be studied by using **sequential-move games**.

Take the case of a new generation of large passenger aircraft that can fly further without refuelling. Assume that there is a market for a 500-seater version of this type of aircraft and a 400-seater version, but that the market for each size of aircraft is not big enough for the two manufacturers, Boeing and Airbus, to share it profitably. Let us also assume that the 400-seater market would give an annual profit of £50 million to a single manufacturer and that the 500-seater would give an

BOX 6.5 **THE PRISONERS' DILEMMA** CASE STUDIES & APPLICATIONS

Choosing whether to deny or confess

Game theory is relevant not just to economics. A famous non-economic example is the prisoners' dilemma.

Carlos and Elina have been arrested for a joint crime of serious fraud. Each is interviewed separately and given the following alternatives:

- First, if they say nothing, the court has enough evidence to sentence both to a year's imprisonment.
- Second, if either Carlos or Elina alone confesses, he or she is likely to get only a three-month sentence but the partner could get up to ten years.
- Third, if both confess, they are likely to get three years each.

These outcomes are illustrated in the diagram. What should Carlos and Elina do?

Let us consider Carlos's dilemma. Should he confess in order to get the short sentence? This is better than the year he would get for not confessing. There is, however, an even better reason for confessing. Suppose Carlos doesn't confess but, unknown to him, Elina does confess. Then Carlos ends up with the long sentence (cell B). Carlos's best strategy is always to confess.

Elina is in the same dilemma and so the result is simple. When both prisoners act in their own self-interest by confessing, they both end up with relatively long prison terms (cell D). Only when they collude will they end up with relatively short ones, the best combined solution (cell A). However, for each of these prisoners, the more certain they are that their compatriot will maintain their innocence, the greater the incentive for them to confess and reduce their sentence!

Of course the police know this and will do their best to prevent any collusion. They will keep Carlos and Elina in separate cells and try to persuade each of them that the other is bound to confess.

Thus the choice of strategy depends on:

- Carlos's and Elina's risk attitudes: i.e. are they 'risk lovers' or 'risk averse'?
- Carlos's and Elina's estimates of how likely the other is to own up.

Alternatives for Carlos and Elina

Carlos's alternatives \ Elina's alternatives	Not confess	Confess
Not confess	A Each gets 1 year	C Carlos gets 10 years Elina gets 3 months
Confess	B Carlos gets 3 months Elina gets 10 years	D Each gets 3 years

1. Why is this a dominant strategy 'game'?
2. How would Carlos's choice of strategy be affected if he had instead been involved in a joint crime with Rikki, Kate, Preethi and Dave, and they had all been caught?

The prisoners' dilemma is a good illustration of **the fallacy of composition**: what applies at the level of the individual does not apply to the group as a whole. It might be in the individual's interests to confess. It is clearly not in the interests of both, however, for both to confess.

KEY IDEA 22 ***The fallacy of composition.*** **What applies in one case will not necessarily apply when repeated in all cases.**

Let us now look at two real-world examples of the prisoners' dilemma.

Standing up to see the action!

Dean is following his beloved Leicester City at yet another match. Leicester break forward at rapid pace. With the prospect of a goal he stands to get a better view. Others follow his lead: after all, if they stayed sitting, they would not see at all.

In this Nash equilibrium, most people are worse off, since, except for tall people, their view is likely to be worse and they lose the comfort of sitting down. If only Dean and others like him could just sit down!

Too much advertising

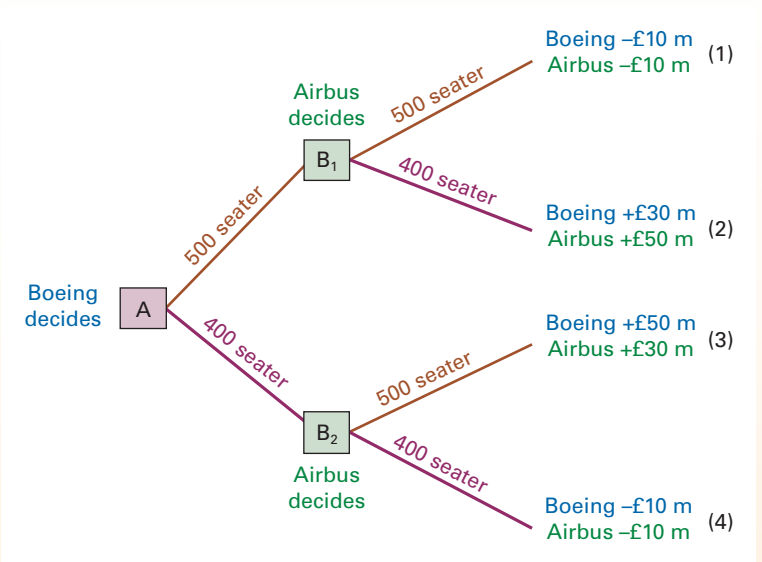
Why do firms spend so much on advertising? If they are aggressive, they do so to get ahead of their rivals. If they are cautious, they do so in case their rivals increase their advertising (the maximin approach).

Although in both cases it may be in the individual firm's best interests to increase advertising, the resulting Nash equilibrium is likely to be one of excessive advertising: the total spent on advertising (by all firms) is not recouped in additional sales.

3. Give one or two other examples (economic or non-economic) of the prisoners' dilemma.

Individually or in small groups arrange a meeting with a manager in a company that clearly competes with rivals. Formulate a series of questions to determine the extent to which the company takes a strategic approach to pricing, advertising and other forms of marketing and the ways in which it (a) tries to anticipate what its rivals will do; (b) how it responds to its rivals' actions; (c) how its rivals respond to its actions. Write a brief report on your findings. It might be appropriate to present this in class.

Figure 6.15 A decision tree



annual profit of £30 million, but that if both manufacturers produced the same version, they would each make an annual loss of £10 million.

Assume that Boeing is the first mover and announces which plane it will build: the 500-seater or 400-seater plane. What should Airbus do? The choice is illustrated in Figure 6.15. This diagram is called a **decision tree** and shows the sequence of events.

The small square at the left of the diagram is Boeing's decision point (point A). If it decides to build the 500-seater plane, we would move up the top branch. Airbus would now have to make a decision (point B_1). If it too built the 500-seater plane, we would move to outcome 1: a loss of £10m for both manufacturers.

Clearly, with Boeing building a 500-seater plane, Airbus would choose the 400-seater plane: we would move to outcome 2, with Boeing making a profit of £30m and Airbus a profit £50m. Airbus would be very pleased!

Boeing's best strategy at point A, however, would be to build the 400-seater plane. We would then move to Airbus's decision point B_2. In this case, it is in Airbus's interests to build the 500-seater plane. Its profit would be only £30m (outcome 3), but this is better than the £10m loss if it too built the 400-seater plane (outcome 4). With Boeing deciding first, the Nash equilibrium will thus be outcome 3.

KI 21 p 140

There is clearly a **first-mover advantage** here. Once Boeing has decided to build the more profitable version of the plane (the 400-seater), Airbus's best response is to build the less profitable one. Naturally, Airbus would like to build the more profitable one and be the first mover. Which company succeeds in going first depends on how advanced they are in their research and development and in their production capacity.

More complex decision trees. The aircraft example is the simplest version of a decision tree, with just two companies and each one making only one key decision. In many business situations, much more complex trees could be constructed. The 'game' would be more like chess, with many moves and several options on each move. If there were more than two companies, the decision tree would be more complex still.

Pause for thought

Give an example of decisions that two firms could make in sequence, each one affecting the other's next decision.

Credibility, threats and promises

In sequential-move games firms will respond not only to what firms do, but what they say they will do. In many situations, an oligopolist will make a threat

Definitions

Decision tree (or game tree) A diagram showing the sequence of possible decisions by competitor firms and the outcome of each combination of decisions.

First-mover advantage When a firm gains from being the first to take action.

TC 6 p 24

or promise that it will act in a certain way. As long as the threat or promise is **credible** (i.e. its competitors believe it), the firm can gain and it will influence its rivals' behaviour.

Take the simple situation where a large oil company, such as Esso, states that it will match the price charged by any competitor within a given radius. Assume that competitors believe this 'price promise' but also that Esso will not try to *undercut* their price. In the simple situation where there is only one other petrol station in the area, what price should it charge? Clearly it should charge the price that would maximise its profits, assuming that Esso will charge the *same* price. In the absence of other petrol stations in the area, this is likely to be a relatively high price.

Pause for thought

Assume that there are two major oil companies operating filling stations in an area. The first promises to match the other's prices. The other promises to sell at 1p per litre cheaper than the first. Describe the likely sequence of events in this 'game' and the likely eventual outcome. Could the promise of the second company be seen as credible?

Now assume that there are several petrol stations in the area. What should the company do now? Its best bet is probably to charge the same price as Esso and hope that no other company charges a lower price and forces Esso to cut its price. Assuming that Esso's threat is credible, other companies are likely to reason in a similar way.

Recap

1. Game theory is a way of modelling behaviour in strategic situations where the outcome for an individual or firm depends on the choices made by others. It enables us to examine various strategies that firms can adopt when the outcome of each is not certain.
2. The simplest type of 'game' is a simultaneous single-move or single-period game. Many simultaneous single-period games have predictable outcomes, no matter what assumptions each firm makes about its rivals' behaviour. Such games are known as dominant strategy games.
3. Many other simultaneous games are more complicated and either one or both firms do not have a dominant strategy. The Nash equilibrium is a useful way to predict the most likely outcome in any of these games.
4. If a simultaneous game is repeated, the equilibrium can change. The final result will depend on a number of factors such as whether the end date of the game is known.
5. In sequential-move games, play is passed from one 'player' to the other sequentially. Firms will respond not only to what firms do, but also to what they say they will do. To this end, a firm's threats or promises must be credible if they are to influence rivals' decisions.
6. A firm may gain a strategic advantage over its rivals by being the first one to take action (e.g. launch a new product). A decision tree can be constructed to show the possible sequence of moves in a multiple-move game.

6.7 PRICE DISCRIMINATION

In what situations will firms be able to charge different prices to different consumers? How will we as consumers benefit or lose from the process?

Up to now we have assumed that a firm sells each unit of output for the same price. This is sometimes referred to as *uniform pricing*. However, if a firm implements a uniform pricing policy, then it is missing out on potential profit. Why? Because some customers gain a higher utility from the product and thus have a greater willingness to pay. They would still purchase the good if the price was higher.

To exploit this situation, the firm might be tempted to increase prices to try to capture some of this consumer surplus and convert it into profit. However, it faces a trade-off. A higher price will increase the profit *per transaction*, but some customers will stop buying the product: i.e. the ones who value it less highly.

Definition

Credible threat (or promise) One that is believable to rivals because it is in the threatener's interests to carry it out.

This trade-off could be avoided, however, if the firm could charge a higher price to those customers with a high valuation for the product (i.e. gaining a high utility) and a lower price to those consumers with a lower valuation for the product. Firms can do this by implementing a strategy of **price discrimination**.

If the cost to a firm of supplying different customers does not vary, then price discrimination can be defined in the following way: it is the practice of selling the same or similar products to different customers for different prices.

If the costs of supplying the good to different customers *do* vary, then the previous definition is incomplete. Hence, we identify price discrimination as the practice of selling the same or similar product at different prices and the difference in price cannot be fully accounted for by any difference in the cost of supply.

Economists traditionally distinguish between three different types of price discrimination.

First-degree price discrimination

First-degree price discrimination is the rarest form of price discrimination. It occurs when sellers charge each consumer the maximum price he or she is willing to pay. It is sometimes called 'perfect price discrimination' or 'personalised pricing'. Stallholders in a bazaar may attempt to do this by initially asking a very high price. Then, as haggling takes place and the price is bid down, a price is settled on that represents the most the consumer is willing to pay.

First-degree price discrimination is generally seen as more of a theoretical benchmark than a viable business strategy since there are few circumstances in which sellers can obtain reliable information on the maximum amount people are willing to pay.

Pause for thought

Demonstrate on a market demand curve how successful first-degree price discrimination would eliminate consumer surplus.

Second-degree price discrimination

Second degree price discrimination involves customers being offered a range of different pricing options for the same or similar product by the firm. Consumers are then free to choose whichever option they wish but the lower prices are conditional on factors such as:

- *The quantity of the product purchased*. In order to obtain the good at a lower price the customer has to purchase a certain quantity of the good or service.
- *The use of coupons/vouchers*. To be eligible to purchase the product for a lower price, customers have to produce a voucher or coupon that they have collected: e.g. from a flyer inside a newspaper or from the Internet.
- *When the product is purchased*. For example, some goods are priced at a higher level when they are first released onto the market. Rail fares are higher at peak times than at off-peak times.
- *The version of the product purchased*. Firms can produce different versions of the same core product that have different levels of actual or perceived quality: e.g. value ranges of own-label products sold in supermarkets. This is called *versioning*. One example of versioning is where firms create a lower quality version of its good by removing some features or reducing its performance characteristics. Note that versioning is not *pure* price discrimination because the product is slightly different.

Third-degree price discrimination

The most common type of price discrimination is referred to as **third-degree** price discrimination. This is where a firm charges a different price to different groups of consumers. In doing so, the firm looks to identify some consumer characteristic, trait or attribute that could be used as a basis to split them into different groups. For example, consumers could be divided by age (e.g. adults and children) or by location (e.g. rich areas and poor areas). In other words, groups of consumers are categorised in some way and then charged a price accordingly.

To be successful the characteristic must have three important properties.

- It must be relatively easy for the firm to observe;
- It must provide some indication of the consumer's willingness to pay: i.e. consumers allocated to one group should generally be less price sensitive at any given price than those allocated to another

Definitions

Price discrimination Where a firm sells the same or similar product at different prices and the difference in price cannot be fully accounted for by any differences in the costs of supply.

First-degree price discrimination Where a firm charges each consumer the maximum price they are willing to pay for each unit.

Second-degree price discrimination Where a firm charges customers different prices for the same (or similar) product depending on factors such as the quantity or time purchased.

Third-degree price discrimination Where a firm divides consumers into different groups based on some characteristic that is relatively easy to observe and acceptable to the consumer. The firm then charges a different price to consumers in different groups, but the same price to all the consumers within each group.

Table 6.4 Examples of third-degree price discrimination

Characteristic	Example
Age	16–25 or senior rail card; half-price children's tickets in the cinema.
Gender	'Ladies' night' in a bar or club where men pay the full price for drinks while women can get the same drinks at a discounted price. Although these are illegal in many countries, some clubs and bars flout the law.
Location	Pharmaceutical companies often charge different prices for the same medicine/drug in different countries. Consumers in the USA are often charged more than those from other countries.
Occupation	Apple, Microsoft and Orange provide price discounts to employees of educational institutions.
Business or individual	Publishers of academic journals charge much lower subscription rates to individuals than university libraries.
Past buying behaviour	Firms often charge new customers a lower price than existing customers for the same product or service as an 'introductory offer'.

TC 10 p 49

group – in other words, price elasticity of demand must differ between the groups;

- It must allow consumers to be split into separate markets so that consumers are unable simply to switch from the higher- to the lower-priced market. For example, people should not be able to purchase the product in the lower-priced market and re-sell it in the higher-priced one.

Table 6.4 provides some examples of the types of characteristics that can be used as the basis for third degree price discrimination.

Advantages to the firm

As we have seen, there are various ways in which firms can engage in discriminatory pricing. Case Study 6.25 on the student website discusses alternative types of price discrimination in more detail. The key point here is that price discrimination allows the firm to earn a higher *revenue* from any given level of sales. To demonstrate this, consider Figure 6.16 which represents a firm's demand curve.

If it is to sell 200 units without price discrimination, assume that it must charge a price of P_1 (the profit-maximising price where $MC = MR$). The total revenue it earns is shown by the blue area. If, however, it can practise second or third-degree price discrimination by selling 150 of those 200 units at the higher price of P_2, it will gain the red striped area in addition to the blue area.

You can also see what would happen if the firm could practise first-degree price discrimination: i.e. charge each consumer the maximum price they are prepared to pay for each unit of the product. In this case, its revenue would be the whole area under the demand curve up to the total amount purchased.

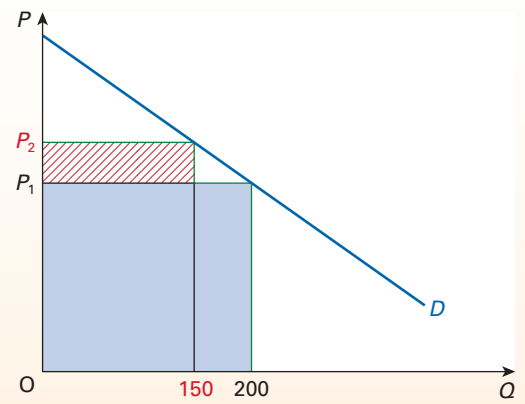
Another advantage the firm gains by price discrimination is that it may be able to use it to drive competitors out of business. If a firm has a monopoly in one market (e.g. the home market), it may be able to charge a high price due to its relatively inelastic demand, and thus make high profits. If it is under oligopoly in another market (e.g. the export market), it may use the high profits in the first market to subsidise a very low price in the oligopolistic market, thus forcing its competitors out of business.

Price discrimination and the public interest

No clear-cut decision can be made over the social desirability of price discrimination. Some people will benefit from it; others will lose. Those paying the higher price will probably feel that price discrimination is unfair to them. On the other hand, those charged the lower price may thereby be able to obtain a good or service they could otherwise not afford: e.g. concessionary bus or train fares for senior citizens.

Misallocation effects. Price discrimination may be allocatively inefficient. Under uniform pricing the product is allocated through the price mechanism to those consumers who value it the most, given their incomes. The implementation of third-degree price discrimination could result in some units

Figure 6.16 Third-degree price discrimination



of the product being re-allocated away from those consumers with a higher willingness to pay to those with a lower willingness to pay.

Without any restrictions, mutually beneficial trades might be able to take place between the buyers. Those consumers with a higher valuation of the good could, under some circumstances, purchase it from those with a lower valuation at a price that would improve the welfare of both parties. However, the seller blocks this re-sale from taking place and in the process reduces society's welfare.

Competition. A firm may use price discrimination to drive competitors out of business. This is known as **predatory pricing**. For example, in many towns, large bus companies have used profits they make in *other* towns where they have a monopoly to subsidise their bus fares and thereby drive competitors out of business, only then to raise prices above those that the competitors had been charging.

KI 19 p 122

On the other hand, it might use its profits from its high-priced market to break into another market and withstand a possible price war. Competition is thereby increased.

Profits. Price discrimination raises a firm's profits. This could be seen as an undesirable redistribution of income in society, especially if the average price of the product is raised. On the other hand, the higher profits may be reinvested and lead to lower costs in the future.

Definition

Predatory pricing Selling at a price below average variable cost in order to drive competitors from the market.

Recap

1. If the cost to the firm of supplying different customers does not vary, then price discrimination is the practice of selling the same or similar products to different customers for different prices. If the cost of supplying different customers does vary, then price discrimination is the practice of selling the same or similar product at different prices where the difference in price cannot be fully accounted for by the difference in the cost of supply.
2. First-degree price discrimination is where a consumer is charged the maximum he or she is prepared to pay. Second-degree price discrimination is where consumers are offered various pricing options for the same or similar product. They are free to choose whichever option they wish but the lower prices are conditional on some features of the purchase. Third-degree price discrimination is where consumers are divided into groups and the groups with the lower price elasticity of demand are charged the higher prices.
3. Price discrimination allows the firm to earn a higher revenue from a given level of sales.
4. Some people will gain from price discrimination; others will lose. It is likely to be particularly harmful when it is used as a means of driving competitors from the market (predatory pricing).

BOX 6.6 **PROFIT-MAXIMISING PRICES AND OUTPUT FOR A THIRD-DEGREE PRICE DISCRIMINATING FIRM** **EXPLORING ECONOMICS**

Identifying different prices in different markets

Assuming that a firm wishes to maximise profits, what discriminatory prices should it charge and how much should it produce?

Assume that the firm sells an identical product in two separate markets H and L with demand and *MR* curves in each market as shown in panels (a) and (b) of the diagram – a case of third-degree price discrimination.

Equilibrium for a single-price firm. If the firm were unable to split its customers into these two different groups, then a market demand curve could be derived. This is illustrated in panel (c) and is obtained by horizontally aggregating the demand curves in panels (a) and (b).

The market demand curve is the same as the demand curve in market H from point *g* to *h*. This is because no consumers in market L are willing to pay a price above *d*. As the price falls below *d* both consumers in market H and L are willing to buy the good, so horizontal aggregation of both demand curves must take place from this point onwards. This creates a kink in the market demand curve at point *h*.

This kink also creates a discontinuity in the *MR* curve between points *j* and *k*. To simplify the explanation, it is also assumed that the firm's marginal cost is constant and that it has no fixed costs. Thus $AC = MC$.

To maximise profit, it would produce where the market marginal revenue equals marginal cost: i.e. $MR_M = MC$. This occurs at point *l* in panel (c). It would therefore produce an output of Q^* and sell all of this output at the same price of P^*.

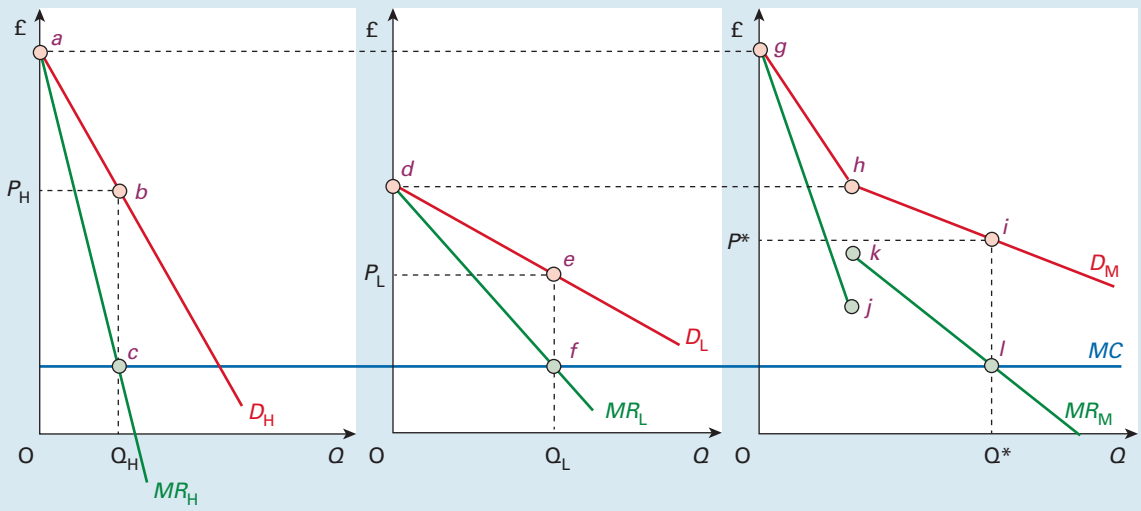
Equilibrium under third-degree price discrimination. What happens if the firm could now charge a different price to the customers in market H from those in market L? At the single price of P^* the price elasticity of demand in market H is lower than it is in market L. (Note that demand is nevertheless elastic in both markets at this price as *MR* is positive.) Therefore the firm could increase its profits by charging a price above P^* in market H and below P^* in market L. This is illustrated in panels (a) and (b).

In market H the profit-maximising firm should produce where $MR_H = MC$. Therefore it should sell an output of Q_H for a price of P_H. In market L it should produce where $MR_L = MC$ at point *f*. Therefore it should sell an output of Q_L for a price of P_L. Note that P_L is below P^*, while P_H is above P^*.

1. *Why is the higher price charged in the market with the less elastic demand curve?*
2. *How easy do you think it would for a firm to split customers into different groups based on their incomes?*

Compile your own list of examples of third-degree price discrimination. In each case identify the characteristics being used to help differentiate between market segments.

Profit-maximising output under third-degree price discrimination



(a) Market H **(b)** Market L **(c)** Total market (markets H + L)

QUESTIONS

1. A perfectly competitive firm faces a price of £14 per unit. It has the short-run cost schedule shown in the table below.

Output	0	1	2	3	4	5	6	7	8
TC (£)	10	18	24	30	38	50	66	91	120

   a. Copy the table and put in additional rows for average cost and marginal cost at each level of output. (Enter the figures for marginal cost in the space between each column.)
   b. Plot *AC, MC* and *MR* on a diagram.
   c. Mark the profit-maximising output.
   d. How much (supernormal) profit is made at this output?
   e. What would happen to the price in the long run if this firm were typical of others in the industry? Why would we need to know information about long-run average cost in order to give a precise answer to this question?
2. If the industry under perfect competition faces a downward-sloping demand curve, why does an individual firm face a horizontal demand curve?
3. On a diagram similar to Figure 6.3, show the long-run equilibrium for both firm and industry under perfect competition. Now assume that the demand for the product falls. Show the short-run and long-run effects.
4. If supernormal profits are competed away under perfect competition, why will firms have an incentive to become more efficient?
5. Is it a valid criticism of perfect competition to argue that it is incompatible with economies of scale?
6. As an illustration of the difficulty in identifying monopolies, try to decide which of the following are monopolies: BT; your local evening newspaper; a water company; the village post office; Interflora; the London Underground; ice creams in the cinema; Guinness; food sold on a train; the board game 'Monopoly®'.
7. Try this brain teaser. A monopoly would be expected to face an inelastic demand. After all, there are no direct substitutes. And yet, if it produces where $MR = MC$, *MR* must be positive, and demand must therefore be elastic. Therefore the monopolist must face an elastic demand! Can you solve this conundrum?
8. For what reasons would you expect a monopoly to charge (a) a higher price, and (b) a lower price than if the industry were operating under perfect competition?
9. In which of the following industries are exit costs likely to be low: (a) steel production; (b) market gardening; (c) nuclear power generation; (d) specialist financial advisory services; (e) production of fashion dolls; (f) production of a new drug; (g) contract catering; (h) mobile discos; (i) car ferry operators? Are these exit costs dependent on how narrowly the industry is defined?
10. Think of three examples of monopolies (local or national) and consider how contestable their markets are.
11. Think of 10 different products or services and estimate roughly how many firms there are in the market. You will need to decide whether 'the market' is a local one, a national one or an international one. In what ways do the firms compete in each of the cases you have identified?
12. Assume that a monopolistically competitive industry is in long-run equilibrium. On a diagram like Figure 6.7, show the effect of a fall in demand on a firm's price and profit in (a) the short run and (b) the long run.
13. Imagine there are two types of potential customer for jam sold by a small food shop. The one is the person who has just run out and wants some now. The other is the person who looks in the cupboard, sees that the pot of jam is less than half full and thinks, 'I will soon need some more'. How will the price elasticity of demand differ between these two customers?
14. Why may a food shop charge higher prices than supermarkets for 'essential items' and yet very similar prices for delicatessen items?
15. Will competition between oligopolists always reduce total industry profits?
16. In which of the following industries is collusion likely to occur: bricks, beer, margarine, cement, crisps, washing powder, carpets?
17. Devise a box diagram like that in Table 6.2 (on page 150), only this time assume that there are three firms each considering the two strategies of keeping price the same or reducing it by a set amount. Is the game still a 'dominant strategy game'?
18. What do you understand by a 'repeated game'? Give examples of how the outcomes from a repeated game may differ from those in single-move games?
19. Is it the size of the purchasing firm that is important in determining its power to keep down the prices charged by its suppliers?
20. If a cinema could sell all its seats to adults in the evenings at the end of the week, but only a few on Mondays and Tuesdays, what price discrimination policy would you recommend to the cinema in order for it to maximise its weekly revenue?
21. Think of two examples of price discrimination. In what ways do the consumers gain or lose? What information would you need to be certain in your answer?
22. Explain why the rise in e-commerce might increase the number of firms engaging in personalised pricing.
23. For a firm to be able to implement a strategy of price discrimination it must be able to prevent re-sale among its customers. What factors would make it more difficult for a consumer to resell a good?

Pearson | MyLab | Economics

Log on to MyLab Economics and complete the studyplan exercises for this chapter to see how much you have learnt and where you need to revise most. Make sure you access all the supporting textbook resources, including the online workbook, newsblog, audio animations, guided solutions and ebook.

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

6.1 **Is perfect best?** An examination of the meaning of the word 'perfect' in perfect competition.

6.2 **B2B electronic marketplaces.** This case study examines the growth of firms trading with each other over the Internet (business-to-business or 'B2B') and considers the effects on competition.

6.3 **Concentration ratios.** One way of measuring the degree of market power in an industry.

6.4 **Windows cleaning.** The examination of Microsoft's market dominance by the US Justice Department.

6.5 **Google.** This case looks at Google's rise in the global Internet search engine market and the EU's investigation of its position in this market.

6.6 **Airline deregulation in the USA and Europe.** Whether the deregulation of various routes has led to more competition and lower prices.

6.7 **Competition in the pipeline?** Monopoly in the supply of gas.

6.8 **X-inefficiency.** A type of inefficiency suffered by many large firms, resulting in a wasteful use of resources.

6.9 **Edward Chamberlin and Joan Robinson.** The birth of the monopolistic competition model.

6.10 **The motor vehicle repair and servicing industry.** A case study of monopolistic competition.

6.11 **Curry wars.** Monopolistic competition in the take-away food market.

6.12 **Bakeries: oligopoly or monopolistic competition?** A case study on the bread industry, showing that small-scale local bakeries can exist alongside giant national bakeries.

6.13 **Oligopoly in the brewing industry.** A case study showing how the UK brewing industry is becoming more concentrated.

6.14 **Supermarket wars?** Is there genuine competition or tacit collusion in the UK supermarket sector?

6.15 **Fixing the price of car parts.** Investigations of global cartels in the car parts industry by competition authorities around the world.

6.16 **Cut throat competition.** An examination of the barriers of entry to the UK razor market.

6.17 **Energising competition in the UK energy sector.** An examination of the UK energy sector which has, in recent times, been dominated by six big firms – the 'Big Six'.

6.18 **Fixing prices of envelopes at mini-golf meetings.** The European Commission's investigation into the market for both standardised and customised paper envelopes in the EU.

6.19 **Competition on the buses?** An examination of the impact of the deregulation of UK bus services in the mid 1980s.

6.20 **A lift to profits?** The EC imposes a record fine on four companies operating a lift and escalator cartel.

6.21 **Merger activity.** This examines mergers in Europe: their causes and consequences.

6.22 **A product's life cycle.** How market conditions vary at different stages in a product's life.

6.23 **Peak load pricing.** An example of price discrimination.

6.24 **How do UK companies set prices?** A summary of the findings of a Bank of England survey on how firms set prices in practice.

6.25 **First, second and third.** An examination of how economists classify different types of price discrimination.

6.26 **Easy pricing.** How low-cost airlines, such as easyJet and Ryanair, use price discrimination to increase their revenue.

WEB APPENDICES

6.1 **Measuring monopoly power.** An examination of how the degree of monopoly power possessed by a firm can be measured.

6.2 **The Cournot model.** An extension of the analysis given in the text.

Chapter 7

Wages and the distribution of income

The large rise in food and fuel prices following the COVID-19 pandemic and the Russian invasion of Ukraine hit the poor particularly badly. It highlighted the large divergences in earnings, with the rich seemingly growing richer, while others struggle to make ends meet. Some see this as symbolic of a crisis in capitalism. However, are such claims justified?

In this chapter, we consider just why it is that some bankers, pop stars and footballers earn such large incomes. Why, on the other hand, do cleaners, hospital porters and workers in clothing factories earn very low incomes?

The explanation for differences in wages lies in the working of labour markets. In the first part of the chapter we will consider how labour markets operate. In particular, we will focus on the determination of wage rates in different types of market: ones where employers are wage takers, ones where they can choose the wage rate, and ones where wage rates are determined by a process of collective bargaining.

In Section 7.3 we ask the more general question of why some people are rich and others poor, and consider the degree of inequality in our society: a society that includes the super-rich, with their luxury yachts and their villas abroad, and people living in slum conditions, with not enough to feed and clothe themselves or their children properly; a society where people begging in the streets are an all too familiar sight.

The chapter closes with a consideration of what can be done to reduce inequality. Is the solution to tax the rich very heavily so that the money can be redistributed to the poor? Or might this discourage people from working so hard? Would it be better, then, to focus on benefits and increase the support for the poor?

After studying this chapter, you should be able to answer the following questions:

- How are wage rates determined in a perfect labour market?
- What are the determinants of the demand and supply of labour and their respective elasticities?
- What forms of market power exist in the labour market and what determines the power of employers and labour?
- What effects do powerful employers and trade unions have on wages and employment?
- How can we measure the extent of inequality in income and wealth and how has inequality changed over time?
- What are the causes of inequality?
- What can the government do to reduce inequality?

7.1 WAGE DETERMINATION IN A PERFECT MARKET

Why are some people paid higher wage rates than others?

Perfect labour markets

When looking at the market for labour, it is useful to make a similar distinction to that made in goods markets: the distinction between perfect and imperfect markets. That way we can gain a clearer understanding of the effects of power, or lack of it, in the labour market. Although in practice few labour markets are totally perfect, many do at least approximate to it.

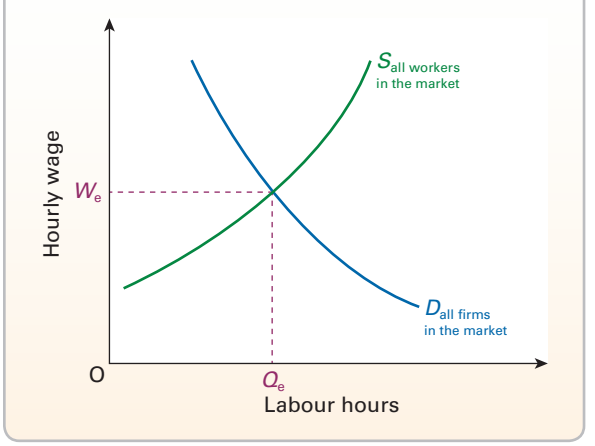
The key assumption of a perfect labour market is that everyone is a **wage taker**. In other words, neither employers nor employees have any economic power to affect wage rates. This situation is not uncommon. Small employers are likely to have to pay the 'going wage rate' to their employees, especially where the employee is of a clear category, such as an electrician, a bar worker, a data analyst or a porter. As far as employees are concerned, being a wage taker means competing with other identical workers and not being a member of a union and therefore not being able to use collective bargaining to push up the wage rate.

Wage rates and employment under perfect competition are determined by the interaction of the market demand and supply of labour. This is illustrated in Figure 7.1. The curves show the total number of hours workers would supply and the number of hours of labour firms would demand for each wage rate in a particular labour market. The equilibrium market wage rate is W_e, where demand equals supply. Equilibrium employment in terms of the total number of hours people are employed in the market is Q_e.

TC 8 p 37

Generally, it would be expected that the supply and demand curves slope the same way as in goods markets.

Figure 7.1 A perfectly competitive labour market



The higher the wage rate paid for a certain type of job, the more workers will want to do that job and, generally, the more hours each will be willing to work. This gives an upward-sloping supply curve of labour. On the other hand, the higher the wage rate that employers have to pay, the less labour they will want to employ. They may simply produce less output, or they may substitute other factors of production, like machinery, for labour. Thus the demand curve for labour slopes downwards.

We now turn to look at the supply and demand for labour in more detail.

The supply of labour

As we have seen, the supply of labour in each market will typically be upward sloping. The *position* of the market supply curve of labour will depend on the number of people willing and able to do the job at each given wage rate. This depends on three things.

- The number of qualified people.
- The non-wage benefits or costs of the job, such as the pleasantness or otherwise of the working environment, job satisfaction or dissatisfaction, status, power, the degree of job security, holidays, perks and other fringe benefits.
- The wages and non-wage benefits in alternative jobs.

A change in the wage rate will cause a movement along the supply curve. A change in any of these other three determinants will shift the whole curve.

TC 5 p 17

Pause for thought

Which way will the supply curve shift if the wage rates in alternative jobs rise?

The elasticity of supply of labour

How *responsive* will the supply of labour be to a change in the wage rate? If the market wage rate goes up, will a lot more labour become available or only a little? This

TC 10 p 49

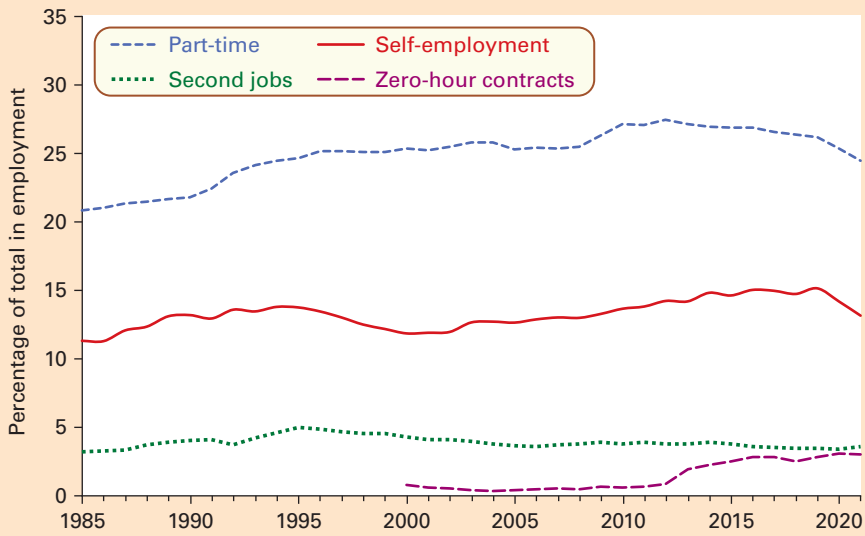
Definition

Wage taker An employer (or employee) who is unable to influence the wage rate.

BOX 7.1 LABOUR MARKET TRENDS

Patterns in employment

Employment trends



Source: Based on data in series MGRZ, YCBH, MGRQ, YCBW and EMP17 (ONS)

The UK labour market has undergone great change in recent years. Advances in technology, changes in the pattern of output, a need to be competitive in international markets and more people working from home – a trend hastened by the pandemic – have contributed to changes in work practices and in the structure and composition of the workforce. Major changes include the following:

Structural change. The UK has experienced a long-term decline in employment in agriculture, forestry and fishing. From around 27 per cent of jobs in the 1860s, agriculture's share of total jobs had fallen to 13 per cent by 1900, 7 per cent by 1930, 5 per cent by 1950 and by 2022 to 1.0 per cent.

The manufacturing sector too has seen a decline in its share of jobs, but its decline began from the 1960s when the sector generated around 30 per cent of jobs. By 1990 this had almost halved to 17 per cent and by 2022 to just 7.2 per cent.

By contrast, jobs in the service industries have grown steadily since the Second World War. In 1950, 53 per cent of jobs were in the service sector. By 1990 this had risen to 72 per cent and to over 84 per cent by 2022.

A rise in female participation rates. By 2022 women constituted nearly 48 per cent of the paid labour force compared with 37 per cent at the beginning of the 1970s. The rise in participation rates is strongly associated with the growth

responsiveness (elasticity) depends on (a) the difficulties and costs of changing jobs and (b) the time period.

Another way of looking at the elasticity of supply of labour is in terms of the **mobility of labour**: the willingness and ability of labour to move to another job, whether in a different location (geographical mobility) or in a different industry (occupational mobility). The mobility of labour (and hence the elasticity of supply of labour) will be higher when there are alternative jobs in the same location, when alternative jobs require similar skills and when people have good information about these jobs.

It is also much higher in the long run, when people have the time to acquire new skills and when the education system has had time to adapt to the changing demands of industry.

The demand for labour: the marginal productivity theory

The market demand curve for labour will typically be downward sloping. To see why, let us examine how many workers an individual firm will want to employ.

In the traditional theory of the firm, which we examined in the previous two chapters, it is assumed that firms aim to maximise profits. The theory of

Definition

Mobility of labour The willingness and ability of labour to move to another job.

CASE STUDIES & APPLICATIONS

in the service sector and the creation of part-time positions (see below). This helps to explain why men still supply a significant majority of the total number of hours worked in the economy. In 2022, men supplied 58 per cent of hours worked, albeit lower than the average of 71 per cent seen across the 1970s.

A rise in part-time employment. In the early 1970s approximately one in six workers in the UK was part time. This figure had risen to over one-fifth of workers by the early 1990s and, as the chart shows, part-time employment had risen to around 27 per cent of workers in the mid-2010s, before falling slightly by the start of the 2020s.

The growth in part-time work reflects the growth in the service sector, where many jobs are part time. These jobs are overwhelmingly held by women. At the start of the 2020s around 75 per cent of part-time jobs have typically been held by women; however, this share has fallen, having been close to 85 per cent in the early 1990s.

Growth in workers with second jobs. The UK has also seen considerable growth in the number of workers with more than one job. Some 1.12 million employees (4.2 per cent of those employed) had second jobs in 2017, of whom 57 per cent were female. This compares with 788 000 (3.2 per cent of those employed) in 1985, of whom 46 per cent were female. The number of people holding second jobs has, however, fallen in recent years.

Growth in self-employed workers. In 2019, 4.97 million of those in employment were classified as self-employed, an historic high. This meant that 15.2 per cent of workers were now self-employed as compared with 11.3 per cent in 1985. Self-employment fell somewhat, however, with the pandemic, as lockdowns meant that many people struggled to maintain their businesses, especially in the hospitality and leisure sectors. Self-employed workers tend to be older than the overall workforce. In 2022 around 45 per cent of self-employed workers were aged 50 or over as compared with 28 per cent of employees. Around one-quarter of those self-employed were in 'skilled trades', such as construction, carpentry and joinery.

A rise in the proportion of workers employed on fixed-term contracts, or on a temporary or casual basis. Many firms nowadays prefer to employ only their core workers/managers on a permanent ('continuing') basis. They feel that it gives them more flexibility in responding to changing market conditions to have the remainder of their workers employed on a short-term basis, and, perhaps, to make use of agency staff or to contract out work (see Case Study 7.2 on the student website).

As part of this trend there has been a growth in zero-hour contracts, where people have no guaranteed hours and are merely offered work as it is available. It was estimated that, at the end of 2021, 1.03 million (3.2 per cent) of those in employment were employed on such contracts (see chart). Of these, 41 per cent were aged 16 to 24, 57 per cent were female and 37 per cent worked in 'elementary occupations'.

The number of workers on zero-hour contracts rose sharply after 2011, when only 190 000 workers or 0.6 per cent of all people in employment were on zero-hour contracts.

What potential issues arise for policy makers as employment patterns change?

Download from the Office for National Statistics dataset A02 SA: Employment, unemployment and economic inactivity for people aged 16 and over and aged from 16 to 64 (seasonally adjusted). Construct two charts for the UK since 1971: total employment by gender and total unemployment by gender. Write a short briefing note to summarise your findings.

labour demand is based on the same assumption. This theory is generally known as the **marginal productivity theory**.

The profit-maximising approach

How many workers should a firm employ in order to maximise profits? The firm will answer this question by weighing up the costs of employing extra labour against the benefits. It will use exactly the same principles as in deciding how much output to produce.

In the goods market, the firm will maximise profits where the marginal cost of producing an extra unit of a *good* equals the marginal revenue from selling it: $MC = MR$. TC 2 p 9

In the labour market, the firm will maximise profits where the marginal cost of employing an extra *worker* equals the marginal revenue that the worker's output earns for the firm: MC of labour $= MR$ of labour. The reasoning is simple. If an extra worker adds more to a firm's revenue than to its costs, the firm's profits will increase. It will be worth employing that worker.

But as more workers are employed, diminishing returns to labour will set in (see pages 94–5). Each extra worker will produce less than the previous one, KI 17 p 95

Definition

Marginal productivity theory The theory that the demand for a factor depends on its marginal revenue product.

and thus earn less revenue for the firm. Eventually the marginal revenue from extra workers will fall to the level of their marginal cost. At that point, the firm will stop employing extra workers. There are no additional profits to be gained. Profits are at a maximum.

Measuring the marginal cost and revenue of labour

Marginal cost of labour (MC_L). This is the extra cost of employing one more worker. Under perfect competition the firm is too small to affect the market wage. It faces a horizontal supply curve. In other words, it can employ as many workers as it chooses at the market wage rate. Thus the additional cost of employing one more person will simply be the wage rate: $MC_L = W$.

Marginal revenue of labour (MRP_L). The marginal revenue that the firm gains from employing one more worker is called the **marginal revenue product of labour** (MRP_L). The MRP_L is found by multiplying two elements – the *marginal physical product* of labour (MPP_L) and the marginal revenue gained by selling one more unit of output (MR).

$$MRP_L = MPP_L \times MR$$

The MPP_L is the extra output produced by the last worker. Thus if the last worker produces 100 tonnes of output per week (MPP_L), and if the firm earns an extra £5 for each additional tonne sold (MR), then the worker's MRP is £500. This extra worker is adding £500 to the firm's revenue.

The profit-maximising level of employment for a firm

The MRP_L curve is illustrated in Figure 7.2. As more workers are employed, there will come a point when diminishing returns set in (point x). Thereafter the MRP_L curve slopes downwards. The figure also shows the MC_L 'curve' at the current market wage W_m.

KI 17 p 95

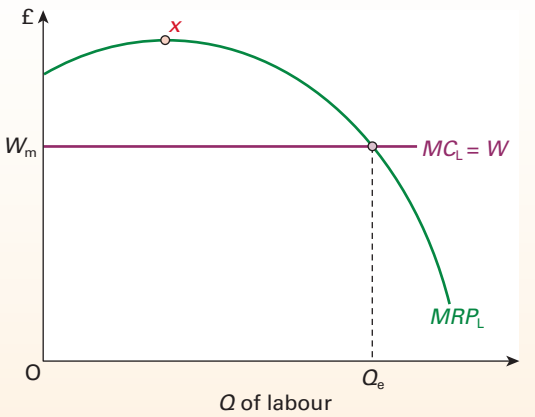
Profits are maximised at an employment level of Q_e, where MC_L (i.e. the wage rate, W) $= MRP_L$. Why? At levels of employment less than Q_e, MRP_L exceeds MC_L. The firm will increase profits by employing more labour. At levels of employment greater than Q_e, MC_L exceeds MRP_L. In this case the firm will increase profits by reducing employment.

TC 2 p 9

Definition

Marginal revenue product (of a factor) The extra revenue a firm earns from employing one more unit of a variable factor: $MRP_{factor} = MPP_{factor} \times MR_{good}$

Figure 7.2 The profit-maximising level of employment



Derivation of the firm's demand curve for labour

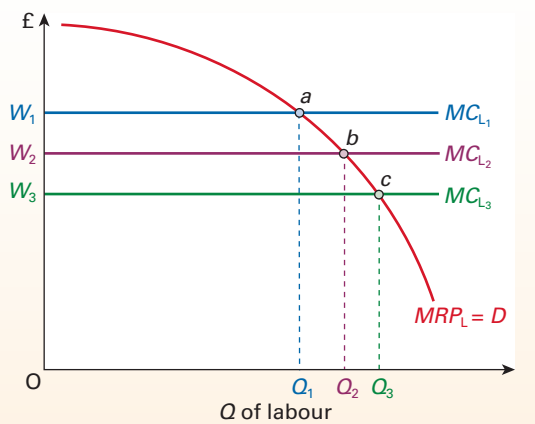
No matter what the wage rate, the quantity of labour demanded is found from the intersection of W and MRP_L (see Figure 7.3). At a wage rate of W_1, Q_1 labour is demanded (point a); at W_2, Q_2 is demanded (point b); at W_3, Q_3 is demanded (point c).

Thus the MRP_L curve shows the quantity of labour employed at each wage rate. But this is just what the demand curve for labour shows. Thus the MRP_L curve is the demand curve for labour.

There are three determinants of the demand for labour by a firm.

- *The wage rate.* This determines the position *on* the demand curve, i.e. the quantity demanded.

Figure 7.3 Deriving the firm's demand curve for labour



- *The productivity of labour* (MPP_L). This determines the position *of* the demand curve.
- *The demand for the good being produced.* The higher the demand for the good, the higher will be its price, and hence the higher will be the *MR*, and thus the MRP_L. This too determines the position *of* the demand curve. It shows how the demand for labour (and other factors) is a **derived demand**: i.e. one derived from the demand for the good. For example, the higher the demand for houses, and hence the higher their price, the higher will be the demand for bricklayers.

Pause for thought

If the productivity of a group of workers rises by 10 per cent, will the wage rate they are paid also rise by 10 per cent? Explain why or why not.

A change in the wage rate is represented by a movement *along* the demand curve for labour. A change in the productivity of labour or in the demand for the good *shifts* the curve.

Market demand for labour and its elasticity

For the same reason that the firm's demand for labour is downward sloping, so the whole market demand for labour will be downward sloping. At higher wage rates, firms in total will employ less labour.

The *elasticity* of this market demand for labour (with respect to changes in the wage rate) depends on various factors. Elasticity will be greater:

TC 10 p 49

The greater the price elasticity of demand for the good. A rise in the wage rate, being a cost of production, will drive up the price of the good. If the market demand for the good is elastic, this rise in price will lead to a lot less being sold and hence a lot fewer people being employed.

The easier it is to substitute labour for other factors and vice versa. If labour can be readily replaced by other inputs (e.g. machinery), then a rise in the wage rate will lead to a large reduction in labour as workers are replaced by these other inputs.

Definition

Derived demand The demand for a factor of production depends on the demand for the good that uses it.

The greater the wage cost as a proportion of total costs. If wages are a large proportion of total costs and the wage rate rises, total costs will rise significantly; therefore production will fall significantly, and so too will the demand for labour.

The longer the time period. Given sufficient time, firms can respond to a rise in wage rates by reorganising their production processes. For example, they could make greater use of automated processes and alternative technologies.

Wages and profits under perfect competition

The wage rate (*W*) is determined by the interaction of demand and supply in the labour market. This will be equal to the value of the output that the last person produces (MRP_L).

Profits to the individual firm arise from the fact that the MRP_L curve slopes downward (diminishing returns), with the last worker adding less to the revenue of firms than previous workers already employed.

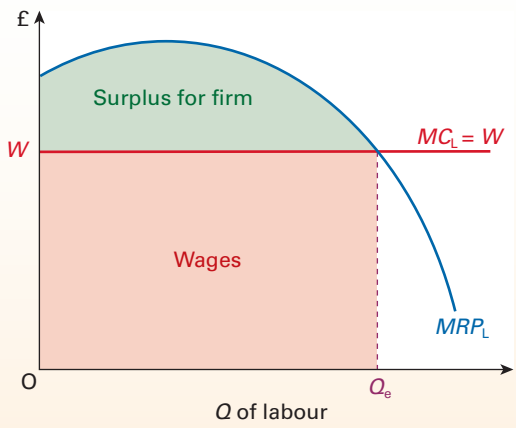
If *all* workers in the firm receive a wage equal to the *MRP* of the last worker, everyone but the last worker will receive a wage *less* than their *MRP*. This excess of MRP_L over *W* of previous workers provides a surplus to the firm over its wages bill (see Figure 7.4). Part of this will be required for paying non-wage costs; part will be profits for the firm.

TC 6 p 24

Perfect competition between firms will ensure that profits are kept down to *normal* profits. If the surplus over wages is such that *supernormal* profits are made, new firms will enter the industry. The price of the good (and hence MRP_L) will fall, and the wage rate will be bid up, until only normal profits remain.

KI 20 p 127

Figure 7.4 Wages and a firm's surplus over wages



Recap

1. Wage rates in a competitive labour market are determined by the interaction of demand and supply.
2. The market supply of the labour curve will normally be upward sloping. Its elasticity will depend on the occupational and geographical mobility of labour. The more readily labour can transfer between jobs and regions, the more elastic will be the supply of labour.
3. The demand for labour depends on labour's productivity. A profit-maximising employer will continue taking on extra workers until MRP_L has fallen (due to diminishing returns) to equal MC_L ($= W$ under perfect competition).
4. The firm's demand curve for labour is its MRP_L curve.
5. The elasticity of demand for labour depends on the elasticity of demand for the good, the ease of substituting labour for other factors and vice versa, wages as a proportion of total costs, and the time period involved.
6. If supernormal profits are made, new firms will enter the industry. The price of the good will fall, and the wage rate will be bid up, until only normal profits remain.

7.2 WAGE DETERMINATION IN IMPERFECT MARKETS

How are wage rates affected by big business and by unions?

Firms with power

In the real world, many firms have the power to influence wage rates: they are not wage takers. This is one of the major types of labour market 'imperfection'.

When a firm is the only employer of a particular type of labour, this situation is called a **monopsony**. Royal Mail used to be a monopsony employer of postal workers.[1] Another example is when a factory is the only employer of certain types of labour in that district. It therefore has local monopsony power. When there are just a few employers, this is called **oligopsony**.

KI 19 p 122

Monopsonists (and oligopsonists too) are 'wage setters' not 'wage takers'. Thus a large employer in a small town may have considerable power to resist wage increases or even to force wage rates down. The National Health Service has considerable power in setting wages for health workers in the UK.

Such firms face an upward-sloping supply curve of labour. This is illustrated in Figure 7.5. If the firm wants to take on more labour, it will have to pay a higher wage rate to attract workers away from other industries. But conversely, by employing less labour it can get away with paying a lower wage rate.

The supply curve shows the wage rate that must be paid to attract a given quantity of labour. The wage rate it pays is the *average cost* to the firm of employing labour (AC_L): i.e. the cost per worker. The supply curve is also therefore the AC_L curve.

The *marginal* cost of employing one more worker (MC_L) will be above the wage (AC_L): see Figure 7.5. The reason is that the wage rate has to be raised to attract extra workers.

The MC_L will thus be the new higher wage paid to the new employee *plus* the small rise in the total wages bill for existing employees: after all, they will be paid the higher wage too.

The profit-maximising employment of labour would be at Q_1, where $MC_L = MRP_L$. The wage (found from the AC_L curve) would thus be W_1.

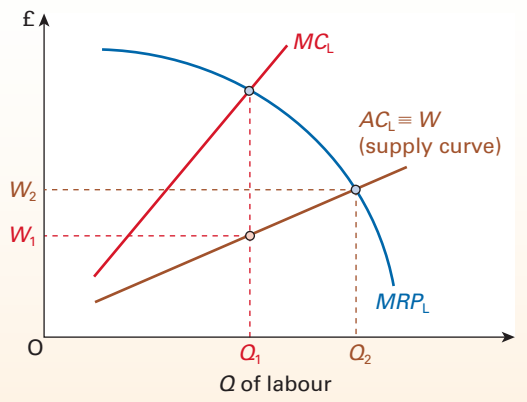
If this had been a perfectly competitive labour market, employment would have been at the higher level Q_2, with

Definitions

Monopsony A market with a single buyer or employer.

Oligopsony A market with just a few buyers or employers.

Figure 7.5 Monopsony



[1] Until 2005, Royal Mail had a statutory monopoly in the delivery of letters.

the wage rate at the higher level W_2, where $W = MRP_L$. The monopsonist is therefore forcing the wage rate down by restricting the number of workers employed.

The role of trade unions

How can unions influence the determination of wages, and what might be the consequences of their actions?

The extent to which unions will succeed in pushing up wage rates depends on their power and willingness to take action. It also depends on the power of firms to resist and on their ability to pay higher wages. In particular, the scope for unions to gain a better deal for their members depends on the sort of market in which the employers are producing.

KI 19 p 122

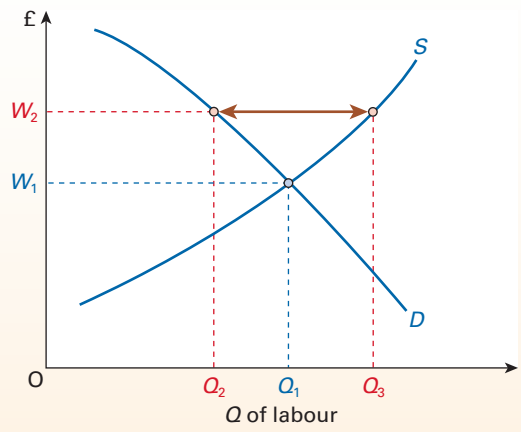
Unions facing competitive employers

If the employers are producing under perfect or monopolistic competition, wage rates can only rise at the expense of employment. Firms are earning just normal profit. Thus if unions force up the wage rate, the marginal firms will make losses and eventually leave the industry. Fewer workers will be employed. The fall in output will lead to higher prices. This will enable the remaining firms to pay a higher wage rate.

Figure 7.6 illustrates these effects. If unions succeed in raising the wage rate from W_1 to W_2, employment will fall from Q_1 to Q_2. There will be a surplus of people ($Q_3 - Q_2$) wishing to work in this industry for whom no jobs are available.

The union faces a second effect. Not only will jobs be lost as a result of the higher wage rate, but also there is a possibility that those in danger of losing their job will leave the union and undercut the union wage (unless employers have an agreement with the union not to employ non-unionised labour).

Figure 7.6 Monopoly union facing producers under perfect competition



In a competitive goods market, wage rates can only be increased without a reduction in the level of employment if, as part of the bargain, the productivity of labour is increased. This is called a **productivity deal**. The *MRP* curve, and hence the *D* curve in Figure 7.6, shifts to the right. Case Study 7.3 on the student website discusses some of the productivity gains from homeworking.

Pause for thought

At what wage rate in Figure 7.6 would employment be maximised: (a) W_1; (b) a wage rate above W_1; (c) a wage rate below W_1? Explain.

Bilateral monopoly

It is common to find the strongest unions where there is a monopsonistic labour market. In these circumstances we can think of the union monopoly as a counterweight to the power of a monopsony employer.

What will the wage rate be under these circumstances? What will the level of employment be? Unfortunately, economic theory cannot give a precise answer. There is no 'equilibrium' level as such (see Box 7.2). Ultimately the wage rate and level of employment will depend on the relative bargaining strengths and skills of unions and management.

Strange as it may seem, unions may be in a better position to make substantial gains for their members when they are facing a powerful employer. There is often considerable scope for them to increase wage rates *without* this leading to a reduction in employment, or even for them to increase both the wage rate *and* employment. The reason is that if firms have power in the *goods* market too, and are making supernormal profit, then there is scope for a powerful union to redistribute some of these profits to wages.

The actual wage rate under bilateral monopoly is usually determined through a process of negotiation or 'collective bargaining'. The outcome of this bargaining will depend on a wide range of factors, which vary substantially from one industry or firm to another.

Collective bargaining

Sometimes when unions and management negotiate, *both* sides can gain from the resulting agreement.

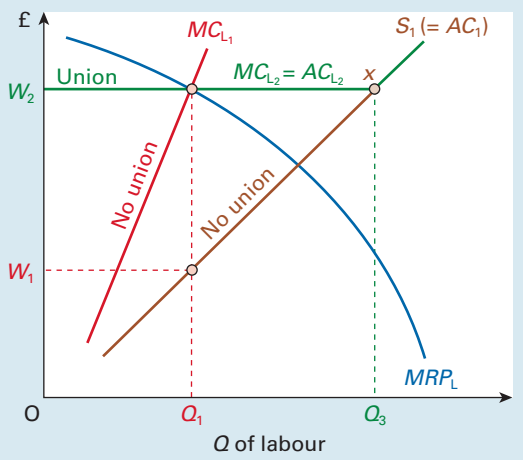
Definition

Productivity deal Where, in return for a wage increase, a union agrees to changes in working practices that will increase output per worker.

BOX 7.2 **WAGES UNDER BILATERAL MONOPOLY** EXPLORING ECONOMICS

All to play for?

Bilateral monopoly



There is no single equilibrium wage rate under bilateral monopoly. This box shows why.

Assume first that there is no union. The diagram shows that a monopsonist employer will maximise profits by employing Q_1 workers at a wage rate of W_1 (Q_1 is where $MRP_L = MC_L$).

What happens when a union is introduced into this situation? Wages will now be set by negotiation between unions and management. Once the wage rate has been agreed, the employer can no longer drive the wage rate down by employing fewer workers. If it tried to pay less than the agreed wage, it could well be faced by a strike, and thus have a zero supply of labour.

Similarly, if the employer decided to take on *more* workers, it would not have to *increase* the wage rate as long as the negotiated wage were above the free-market wage: as long as the wage rate were above that given by the supply curve S_1.

The effect of this is to give a new supply curve that is horizontal up to the point where it meets the original supply curve. For example, let us assume that the union succeeds in negotiating a wage rate of W_2. The supply curve will be horizontal at this level to the left of point *x*. To the right of this point it will follow the original supply curve S_1, since to acquire more than Q_3 workers it would have to raise the wage rate above W_2.

If the supply curve is horizontal to the left of point *x* at a level of W_2, so too will be the MC_L curve. The reason is simply that the extra cost to the employer of taking on an extra worker (up to Q_3) is merely the negotiated wage rate: no rise has to be given to existing employees. If MC_L is equal to the wage, the profit-maximising employment ($MC_L = MRP_L$) will now be where $W = MRP_L$. At a negotiated wage rate of W_2, the firm will therefore choose to employ Q_1 workers.

What this means, therefore, is that the union can push the wage rate up from W_1 to W_2 and the firm will still *want* to employ Q_1. In other words, a wage rise can be obtained *without* a reduction in employment.

The union could go further still. By threatening industrial action, it may be able to push the wage rate above W_2 and still insist that Q_1 workers are employed (i.e. no redundancies). The firm may be prepared to see profits drop right down to normal level rather than face a strike and risk losses. The absolute upper limit to the wage rate will be that at which the firm is forced to close down.

KI 15 p 80

1. *If the negotiated wage rate were somewhere between W_1 and W_2, how would the resulting level of employment compare with that at W_1?*
2. *What in practice will determine just how much the agreed wage rate is above W_1?*

For example, the introduction of new technology may allow higher wages, improved working conditions and higher profits. Usually, however, one side's gain is the other's loss. Higher wages mean lower profits.

The outcome of the negotiations will depend on the relative bargaining strengths of both sides. In bargaining there are various threats or promises that either side can make. For these to be effective, of course, the other side must believe that they will be carried out – that the threats are credible.

Union *threats* might include strike action, **picketing**, **working to rule** or refusing to co-operate with management – for example, in the introduction of new technology. Alternatively, in return for higher wages or better working conditions, unions might *offer* no-strike agreements, increased productivity, or long-term deals over pay.

In turn, employers might *threaten* employees with redundancies or reduced benefits. Alternatively, they might *offer*, in return for lower wage increases, better rewards such as productivity bonuses, profit-sharing schemes, longer holidays or greater job security.

Industrial action imposes costs on both unions and firms. Union members lose pay; firms lose revenue. It is usually in both sides' interests, therefore, to settle

Definitions

Picketing Where people on strike gather at the entrance to the firm and attempt to dissuade workers or delivery vehicles from entering.

Working to rule Where union members are instructed to stick to the letter of their job description and to refuse to take on any extra duties.

by negotiation. Nevertheless, to gain the maximum advantage, each side must persuade the other that it will carry out its threats if pushed.

The approach described so far has essentially been one of confrontation. The alternative is for both sides to concentrate on increasing the total net income of the firm by co-operating on ways to increase efficiency or the quality of the product. This approach is more likely when unions and management have built up an atmosphere of trust over time.

Pause for thought

What parallels are there between union and management strategies and the various strategies that rival oligopolists can adopt?

Union membership. Trade union membership varies significantly around the world. In Finland and Sweden, it is around 70 per cent of employees. In France, by contrast, it is just 11 per cent. In the UK in 2021, 6.72 million people (23.1 per cent of employees) were trade union members.

Membership in many countries has been falling, however. This can be explained by a number of factors: the shift to a service-based economy; continued privatisation and the introduction of private-sector management practices, such as local pay bargaining; and contracted-out services into many of the remaining parts of the public sector. More women working and more part-time and casual work, with more people on 'zero-hour contracts' (see Box 7.1), are also contributory factors, as are the attitudes of many firms to union recognition.

Union membership remains highest in areas of the public sector with high levels of monopsony power, such as education, but there is no doubt that, even here, their power has declined.

Case Study 7.6 on the student website charts the rise and decline of the labour movement in the UK.

The efficiency wage hypothesis

We have seen that a union may be able to force an employer to pay a wage above the market-clearing rate. But it may well be in employers' interests to do so, even in non-unionised sectors.

One explanation for this phenomenon is the **efficiency wage hypothesis**. This states that the productivity of workers rises as the wage rate rises. As a result, employers are frequently prepared to offer wage rates above the market-clearing level, attempting to balance increased wage costs against gains in productivity. But why may higher wage rates lead to higher productivity? There are three main explanations.

KI 20 p 127

Less 'shirking'. In many jobs it is difficult to monitor the effort that individuals put into their work. Workers may thus get away with shirking or careless behaviour. This is an example of the *principal–agent problem* (see Section 5.5).

KI 18 p 117

The business could attempt to reduce shirking by imposing a series of sanctions, the most serious of which would be dismissal. The greater the wage rate currently received, the greater will be the cost to the individual of dismissal, and the less likely it is, therefore, that workers will shirk. The business will benefit not only from the additional output, but also from a reduction in the costs of having to monitor workers' performance.

As a consequence, the **efficiency wage rate** for the business will be the 'true' profit-maximising wage rate for the firm. In this interpretation, the efficiency wage rate is the actual equilibrium wage rate determining the equilibrium level of employment.

Reduced labour turnover. If workers receive on-the-job training or retraining, then to lose a worker once the training has been completed is a significant cost to the business. Labour turnover, and hence its associated costs, can be reduced by paying a wage above the market-clearing rate. By paying such a wage rate, the business is seeking a degree of loyalty from its employees.

Improved morale. A simple reason for offering higher wage rates is to motivate the workforce – to create the feeling that the firm is a 'good' employer that cares about its employees. As a consequence, workers might be more industrious and more willing to accept the introduction of new technology (with the reorganisation that it involves).

TC 6 p 24

The paying of efficiency wages will depend upon the type of work involved. Workers who occupy skilled positions, especially where the business has invested time in their training (thus making them costly to replace), are likely to receive relatively high efficiency wages. By contrast, workers in unskilled positions, where shirking can be easily monitored, where little training takes place and where workers can be easily replaced, are unlikely to command an 'efficiency wage premium'. In such situations, rather than keeping wage rates high, the business will probably try to pay as little as possible.

Definitions

Efficiency wage hypothesis The hypothesis that the productivity of workers is affected by the wage rate that they receive.

Efficiency wage rate The profit-maximising wage rate for the firm after taking into account the effects of wage rates on worker motivation, turnover and recruitment.

Recap

1. Where a firm has monopoly power in employing labour, it is known as a 'monopsonist'. Such a firm will employ workers to the point where $MRP_L = MC_L$. Since the wage is below MC_L, the monopsonist, other things being equal, will employ fewer workers at a lower wage than would be employed in a perfectly competitive labour market.
2. If a union has monopoly power, its power to raise wages will be limited if the employer operates under perfect or monopolistic competition in the goods market. A rise in wage rates will force the employer to cut back on employment, unless there is a corresponding rise in productivity.
3. In a situation of bilateral monopoly (where a monopoly union faces a monopsony employer) the union may have considerable scope to raise wage rates above the monopsony level. There is no unique equilibrium wage. The wage rate will depend on the outcome of a process of collective bargaining between union and management.
4. Collective bargaining is the process by which employers and unions negotiate wage levels and the terms and conditions of employment. Both sides can use threats and promises to determine the outcome of the negotiating process. The success of such threats and promises depends upon factors such as the power of the union or the employer; attitudes and the determination to win; scope for compromise; negotiating skills; information; and the role of government.
5. The efficiency wage hypothesis states that a firm might pay a wage premium to: reduce shirking; reduce labour turnover; improve the quality of labour recruited; and stimulate worker morale. The level of efficiency wage rates will largely be determined by the types of job workers do, and the level and scarcity of the skills they possess.

7.3 INEQUALITY

How do we assess the extent of inequality?

Types of inequality

There a number of different ways of looking at the distribution of income and wealth and so assessing the extent of inequality.

The distribution of income

There are three broad ways of examining the distribution of income.

The size distribution of income. First, we can look at how evenly incomes are distributed among the population. This is known as the **size distribution of income**. It can be expressed between *households*, or between *individual earners*, or between *all individuals*. It can be expressed either *before* or *after* the deduction of taxes and the receipt of benefits. For example, we might want to know the proportion of pre-tax national income going to the richest 10 per cent of households.

The functional distribution of income. We could also look at the distribution between different *factors of production*, known as the **functional distribution of income**. At the *broader* level, we could look at the distribution between the general factor categories: labour, land and capital. At a *narrower* level, we could look at distribution within the factor categories. Why are some jobs well paid while others are badly paid? Why are rents higher in some areas than in others?

Distribution by groups of people. Finally we could look at the **distribution of income by class of recipient**. This can be by *class of person*: women, men, single people, married people, people within a particular age group or ethnic group, and so on. Alternatively, it can be by *geographical area*. Typically, this is expressed in terms of differences in incomes between officially defined regions within a country.

When the distribution of income is analysed, different measures of income can be used. These reflect the different stages of redistribution and, hence, the impact of government intervention. Households' **original income** is made up of employment income, income from financial products and income from

Definitions

Size distribution of income Measurement of the distribution of income according to the levels of income received by individuals (irrespective of source).

Functional distribution of income Measurement of the distribution of income according to the source of income (e.g. from employment, from profit, from rent, etc.).

Distribution of income by class of recipient Measurement of the distribution of income between the classes of person who receive it (e.g. homeowners and non-homeowners or those in the north and those in the south).

Original income Income before taxes and benefits.

Table 7.1 Distribution of UK income by quintile group of households

Household income group	1980		1990		2000/1		2010/11		2020/21	
	Original income	Disposable income	Original income	Disposable income	Original income	Disposable income	Original income	Disposable income	Original income	Disposable income
Lowest 20%	2.1	7.0	1.7	6.0	2.2	6.5	3.2	7.3	3.8	7.0
Next 20%	10.1	13.0	7.0	10.3	7.4	11.3	7.3	11.7	8.9	12.4
Middle 20%	19.3	19.2	16.6	17.3	15.6	16.7	14.1	16.2	15.0	16.7
Next 20%	27.0	24.8	26.5	24.5	25.8	23.6	24.7	22.9	22.6	22.3
Highest 20%	41.5	35.9	48.1	41.9	49.0	41.9	50.8	42.0	49.7	41.5
All	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: Figures may not sum to 100 due to rounding

Source: Based on data from *The effects of taxes and benefits on UK household income, UK, 2020/21*, Reference Table 2 (ONS, 28 March 2022)

occupational pensions. It is income before government intervention.

When we add to original income any cash benefits paid by government, such as state pensions, we have **gross income**. Next, after deducting direct taxes, such as income tax and national insurance contributions we have **disposable income**. If we now deduct indirect taxes, like value added tax (VAT) and other expenditure taxes, we have **post-tax income**. Finally, we can attempt to add on the estimated value of *benefits in kind*, such as health and education services, and, in doing so, obtain **final income**.

Table 7.1 shows the size distribution in the UK for *original* and *disposable* income. The population is divided into five equal-sized groups or 'quintiles'.

The following points can be drawn from these statistics:

- In 2020/21, the richest 20 per cent of households earned 49.7 per cent of original income, and even after the deduction of income taxes and national insurance their share of disposable income was still 41.5 per cent.
- The poorest 20 per cent, by contrast, earned a mere 3.8 per cent of original income, and even after the receipt of cash benefits their share of disposable income was only 7 per cent.

Inequality began to grow dramatically from the 1980s in UK and many other countries. Between 1980 and 1990, the share of disposable income of the poorest 40 per cent of households fell from 20 per cent to 16.3 per cent; while the share of the top 20 per cent grew from 35.9 per cent to 41.9 per cent.

From 2000 to 2010 the income share of the poorest quintiles rose slightly. This reflected the impact of minimum wages and more generous cash benefits. Inequality data show that cash benefits consistently have the largest effect on reducing income inequality. Since 2010, inequality has remained roughly constant.

However, those on the very highest incomes – the top 1 per cent and especially the top 0.1 per cent – have seen an enormous growth in their incomes. Although there was a slight fall in the top income group's share of national income directly following both the financial crisis and the pandemic, the gap then widened again. A similar picture has emerged in many other developed countries.

The distribution of wealth

Income is a *flow*. It measures the receipt of money per period of time (e.g. £25 000 per year). Wealth, by contrast, is a *stock*. It measures the value of a person's assets at a particular point in time.

KI 23 p 173

The distribution of wealth can be measured as a size distribution (how evenly it is distributed among the population); as a functional distribution (the proportion of wealth held in various forms, such as dwellings, land, company shares, bank deposits, etc.); or according to the holders of wealth, classified by age, sex, geographical area, etc.

KEY IDEA 23

Stocks and flows. A stock is a quantity of something at a given point in time. A flow is an increase or decrease in something over a specified period of time. This is an important distinction and a common cause of confusion.

Definitions

Gross income Original income plus cash benefits.

Disposable income Original income plus cash benefits and minus direct taxes and other deductions.

Post-tax income Disposable income minus indirect taxes.

Final income Original income plus the addition of all benefits (cash and in kind) and the deduction of all taxes (direct and indirect).

Measuring the size distribution of income

Two of the most widely used methods for measuring inequality are the *Lorenz curve* and the *Gini coefficient*. We introduce them here in the context of the inequality of income, but they can also be used to measure the inequality of wealth.

Lorenz curve

Figure 7.7 shows a hypothetical **Lorenz curve**. The horizontal axis measures percentages of the population from the poorest to the richest. Thus the 50 per cent point represents the poorest 50 per cent of the population. The vertical axis measures the percentage of national income they receive.

The curve starts at the origin: zero people earn zero incomes. If income were distributed totally equally, the Lorenz curve would be a straight 45° line. The 'poorest' 20 per cent of the population would earn 20 per cent of national income; the 'poorest' 50 per cent would earn 50 per cent, and so on. The curve ends up at the top right-hand corner, with 100 per cent of the population earning 100 per cent of national income.

In practice, the Lorenz curve will 'hang below' the 45° line. Point *x*, for example, shows a country where the poorest 50 per cent of households receive only 25 per cent of national income. The further the curve drops below the 45° line, the greater will be the level of inequality.

The Lorenz curve is quite useful for showing the change in income distribution over time. From 1949 to 1979 the curve for the UK moved inwards towards the 45° line, suggesting a lessening of inequality. Then from 1979 to 1990 it moved downwards away from the 45° line, suggesting a deepening of inequality. Since 1990 it has remained approximately the same.

The problem with simply comparing Lorenz curves by eye is that it is imprecise. This problem is overcome by using Gini coefficients.

Gini coefficient

The **Gini coefficient** is a precise way of measuring the position of the Lorenz curve. It is the ratio of the area between the Lorenz curve and the 45° line to the whole area below the 45° line. In Figure 7.7 this is the ratio of the shaded area *A* to the whole area (*A* + *B*), sometimes expressed as a percentage.

If income is totally equally distributed so that the Lorenz curve follows the 45° line, area *A* disappears and the Gini coefficient is zero. As inequality increases, so does area *A*. The Gini coefficient rises. In the extreme case of total inequality, where one person earns the whole of national income, area *B* would disappear and the Gini coefficient would be 1. Thus the Gini coefficient will be between 0 and 1. The higher it is, the greater is the inequality.

A Gini coefficient can be calculated for each of the different measures of income reflecting the different

Definitions

Lorenz curve A curve showing the proportion of national income (or wealth) earned by any given percentage of the population (measured from the poorest upwards).

Gini coefficient The area between the Lorenz curve and the 45° line divided by the total area under the 45° line.

Figure 7.7 Lorenz curve

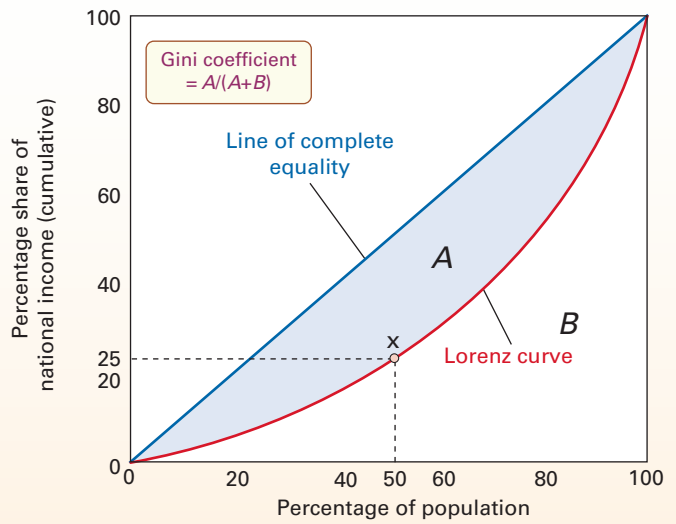
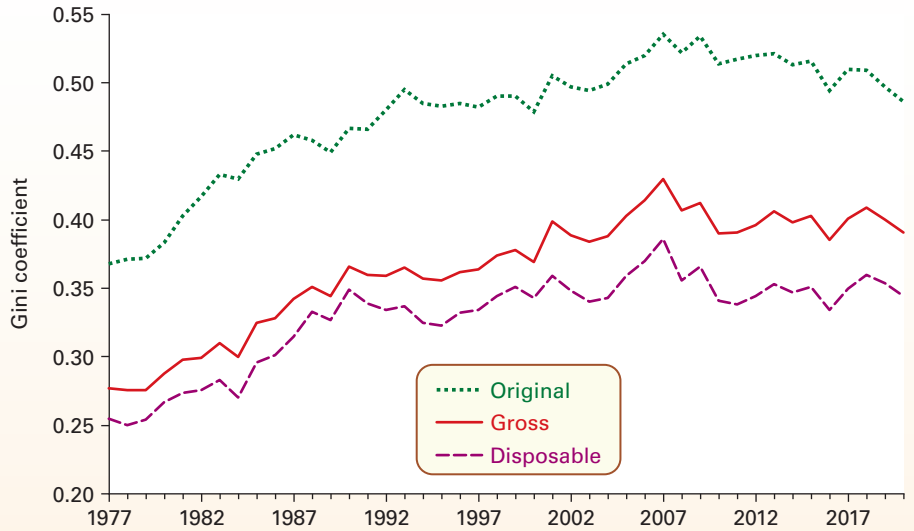


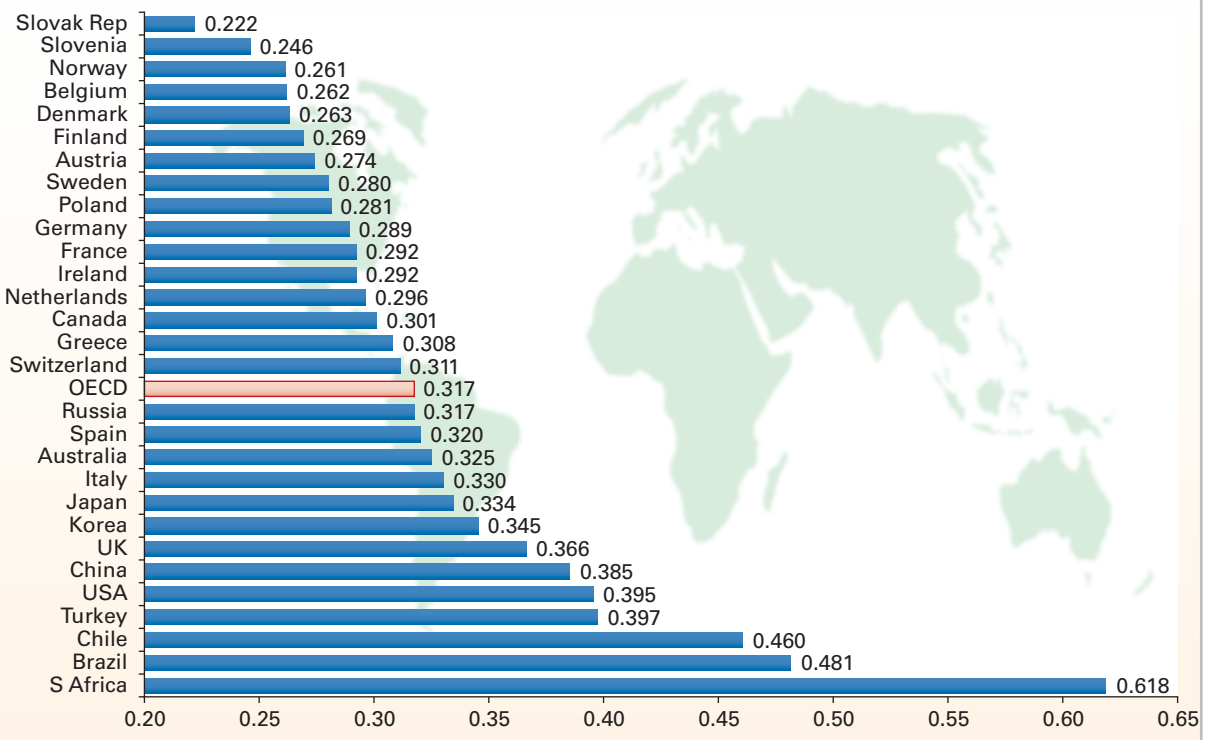
Figure 7.8 UK Gini coefficients: 1977 to 2020/21



Note: From 1994/5 the figures refer to financial years beginning in the year shown

Source: Based on data from *The effects of taxes and benefits on UK household income, UK, 2020/21,* Reference Table 9 (ONS, 28 March 2022)

Figure 7.9 Gini coefficients for selected OECD countries, based on disposable income: latest year, 2016 to 2020



Source: Based on data in *OECD income inequality*

stages of redistribution. The most commonly used is the Gini coefficient for *disposable income*.

In 1977 the *disposable income* Gini coefficient in the UK was 0.255 (see Figure 7.8). With the growth in inequality during the 1980s, the coefficient steadily increased and stood at 0.349 in 1990. Then it gradually fell until the mid-1990s, but rose to 0.359 by 2001/2. It then fluctuated through the 2000s and 2010s, varying from a high of 0.386 in 2007/8 to a low of 0.334 in 2016/17. In 2020/21 it stood at 0.344.

Figure 7.9 shows the Gini coefficients for a selection of countries. These are based on *disposable income*. As you can see, several northern European countries were among those countries with the lowest Gini coefficients and hence were the most equal. The most unequal in the sample was South Africa. The average across the OECD[2] countries was 0.317, more equal than both the USA (0.395) and the UK (0.366).

The functional distribution of income

Distribution of income by source

Figure 7.10 shows the sources of gross household incomes in 1977 and 2020/21. Wages and salaries constitute by far the largest element. However, their share fell from 77 per cent to 68 per cent of gross household incomes over this period. Conversely, the share coming from pensions rose from 2 per cent to 6 per cent, reflecting the growing proportions of the population past retirement age.

In contrast to wages and salaries, investment income (dividends, interest and rent) accounts for a relatively small percentage of household income – a mere 4 per cent in 2020/21. This partly reflects the historically very low rates of interest paid on savings accounts.

The growth of small businesses and the increased numbers of people being 'employed' on a freelance basis has seen the proportion of incomes coming from self-employment increase slightly from around 8 to 9 per cent.

The overall shares illustrated in Figure 7.10 hide the fact that the sources of gross income differ quite markedly between different income groups. These differences are shown in Table 7.2.

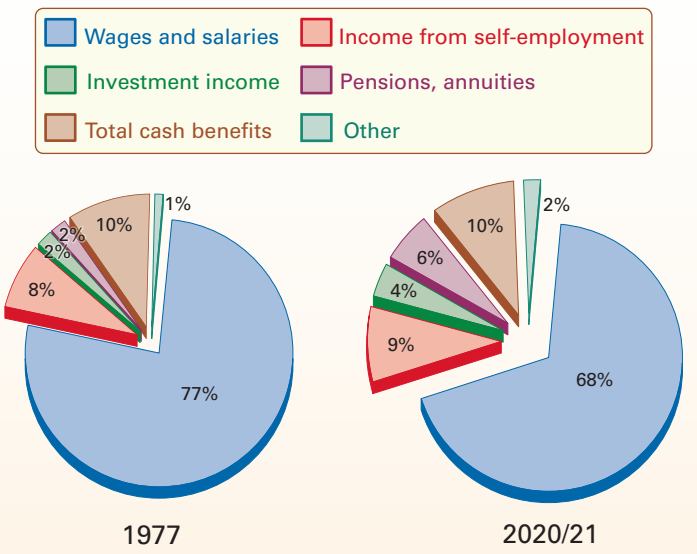
Column (1) shows that higher-income groups get a larger proportion of their income from wages and salaries than do lower-income groups. As would be expected, the poor tend to get a larger proportion of their incomes from social security benefits than do people further up the income scale (see column (5)).

We can see from the table that the proportion of income coming from profits, rent and interest (column (3)) varies little between the income groups. The exception is for the richest income group and even here its significance tends to be concentrated among the very richest. The conclusion from this, plus the fact that investment incomes account for only 4.2 per cent of household incomes in total, is that incomes from capital and land are of only relatively minor significance in explaining income inequality.

Distribution of wages and salaries by occupation

The major cause of differences in incomes between individuals in employment is the differences in wages and salaries between different occupations. Differences in full-time wages and salaries are illustrated in Figure 7.11. This shows the average gross weekly

Figure 7.10 Sources of UK household income as a percentage of gross household income



Source: Based on data in *The effects of taxes and benefits on household income, UK, 2020/21*, Reference Table 28 (ONS, 28 March 2022)

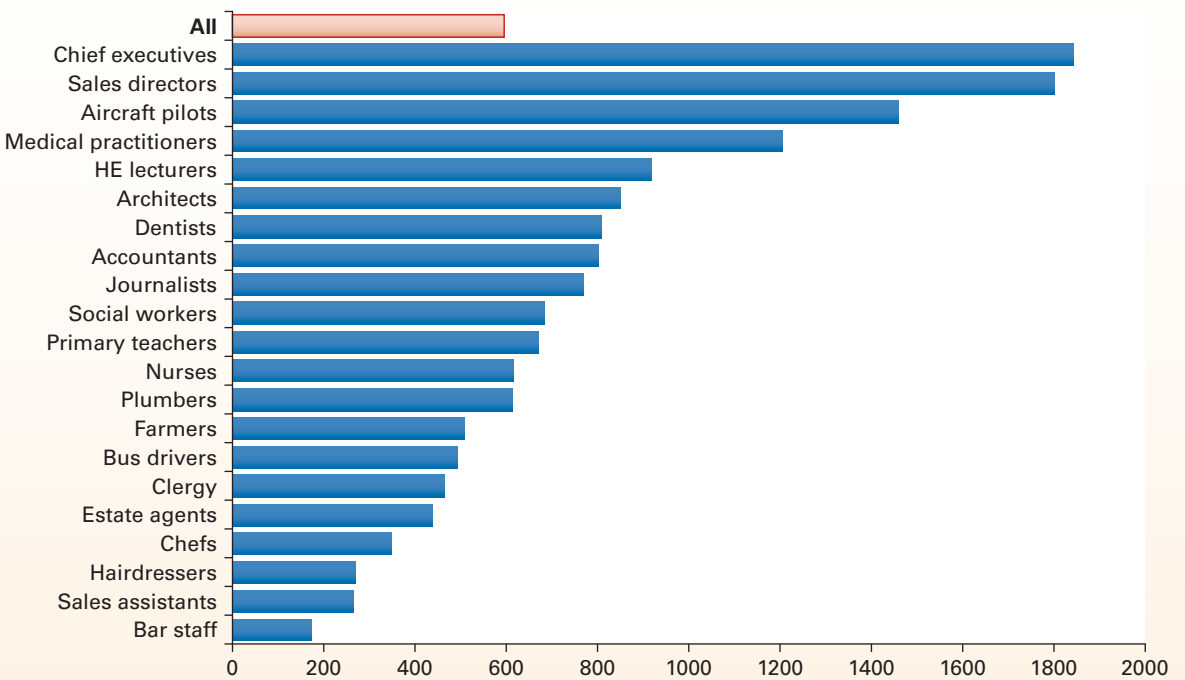
[2]The Organisation for Economic Co-operation and Development (OECD): a group of 38 industrialised countries.

Table 7.2 Sources of UK household income as a percentage of total household gross weekly income by quintile groups: 2020/21

Gross household weekly incomes (quintiles)	Wages and salaries	Income from self-employment	Income from investments	Pensions and annuities	Social security benefits	Other	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lowest 20%	40.5	7.8	1.3	4.7	43.7	1.9	100.0
Next 20%	57.7	6.7	1.4	6.4	25.7	2.1	100.0
Middle 20%	69.7	7.4	1.2	7.1	12.6	1.9	100.0
Next 20%	75.6	6.7	2.2	7.3	6.3	1.9	100.0
Highest 20%	71.3	10.5	7.3	6.0	2.0	2.9	100.0
All households	68.5	8.6	4.2	6.4	10.0	2.3	100.0

Source: Based on data from *The effects of taxes and benefits on household income, UK, 2020/21*, Reference Table 13 (ONS, 28 March 2022)

Figure 7.11 Mean gross weekly earnings (excluding overtime) of UK full-time adult employees (£): 2021



Source: Based on data in the *Annual Survey of Hours and Earnings*, Table 14.2a (National Statistics, October 2021)

earnings of full-time adult workers in selected occupations in the UK in 2021. As you can see, there are considerable differences in pay between different occupations.

Differences in weekly earnings are mainly the result of differences in hourly rates of pay (see Box 7.3), but are also partly explained by differences in the number of hours worked (including the amount of overtime).

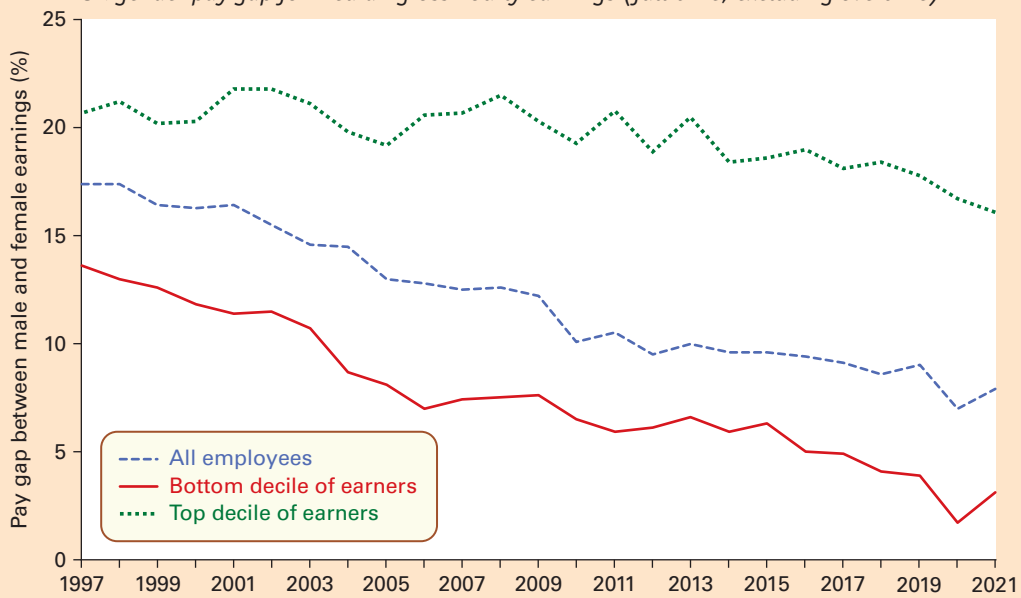
Pause for thought

If fringe benefits (such as long holidays, company cars, free clothing/uniforms, travel allowances and health insurance) were included, do you think the level of inequality would increase or decrease? Explain why.

BOX 7.3 THE GENDER PAY GAP

Wage inequalities between men and women

UK gender pay gap for median gross hourly earnings (full-time, excluding overtime)



Note: Discontinuities in series in 2004, 2006, 2011 and 2018

Source: Based on data in *Gender pay gap in the UK: 2021* (ONS, October 2021)

Women in employment earn less than men. How much less depends on how earnings are measured – whether we include both full-time and part-time employees and the way that earnings are then averaged. The most widely used measure of the gender gap is based on the percentage difference in the *average* gross hourly earnings of employees. Figures in the Annual Survey of Hours and Earnings (ASHE) show that in the UK in 2021, the gender pay gap across all employees was between 14.9 (mean average) and 15.4 per cent (median average).

A similar picture can be seen in the EU. In 2020, women's *mean* average hourly pay was 87.0 per cent of men's for the EU economy. The gender pay gap varies from one country to another. In Luxembourg and Romania it was only 0.7 and 2.4 per cent respectively, whereas in Latvia it was 22.3 per cent and 18.3 per cent in Germany.

Data from the USA indicate that women earn about 84 per cent of their male counterparts. In Japan, the figure is 77.5 per cent.

It is important not to confuse the gender pay gap with unequal pay. Unequal pay refers to situations where men and women receive different rates of pay even though they are doing exactly the same job. This is illegal in the UK under the 2010 Equality Act, although it probably still exists in some workplaces.

A major explanation for the pay gap is *occupational segregation*, with women more likely to be employed in poorly paid occupations. But if you consider the table you can see that quite substantial earnings differentials persist *within* particular occupations.

So what might explain this pay gap? There are a number of possible reasons:

- The marginal productivity of labour in typically female occupations may be lower than in typically male occupations. This may in part be due to simple questions of physical strength. More often, however, it is due to the fact that women tend to work in more labour-intensive occupations. If there is less capital equipment per female worker than there is per male worker, then the marginal product of a woman is likely to be less than that of a man. Evidence from the EU suggests that occupational segregation is a significant factor in explaining pay differences.
- Many women take career breaks to have children. For this reason, employers are sometimes more willing to invest money in training men (thereby increasing their marginal productivity), and more willing to promote men. A study by the Institute for Fiscal Studies (IFS)[1] estimated that for every year a woman takes away from work her earnings fall by 2 per cent below those who remain in work.
- Women tend to be less geographically mobile than men. If social norms are such that the man's job is seen as somehow more 'important' than the woman's, then a couple will often move if that is necessary for the man to get promotion. The woman, however, will have to settle for whatever job she can get in that same locality. This may also reduce a woman's bargaining power when negotiating for wage increases in her current job, if her employer knows that her outside options are more limited than a man's would be.
- While the overall gender gap across part-time workers (mainly women) is much smaller and, under some measures reversed, with men paid less than women, the IFS study noted the longer-term detrimental effect on women's earning when switching from full-time to part-time employment. When this happens, their growth in earnings is likely to fall behind that of people working full time.
- A smaller proportion of female workers are members of unions than male, though the gap is narrowing. Even when they are members of unions, they are often in jobs where unions are weak (e.g. clothing industry workers and shop assistants).
- Custom and practice. Despite equal pay legislation, many jobs done wholly or mainly by women continue to be low paid, irrespective of productivity.

[1] *The Gender Wage Gap*, Institute for Fiscal Studies, August 2016.

Average (mean) hourly pay, excluding overtime, for selected occupations, full-time UK employees on adult rates, 2021

Occupation	Men	Women	Women's pay as % of men's
	£ per hour		
Aircraft pilots and flight engineers	43.19	21.98	50.9
Brokers	37.62	22.60	60.1
Chief executives and senior officials	59.26	42.76	72.2
Actuaries, economists, statisticians	32.03	24.76	77.3
Legal professionals	37.51	30.21	80.5
HE teaching professionals	32.06	26.67	83.2
Librarians	18.52	15.56	84.0
Medical practitioners	34.43	29.43	85.5
Hairdressers and barbers	10.01	8.86	88.5
Laboratory technicians	13.67	12.25	89.6
Chartered and certified accountants	24.32	22.05	90.7
Nurses	21.07	19.28	91.5
Secondary education teaching professionals	25.59	23.46	91.7
Sales assistants and retail cashiers	11.46	10.68	93.2
Management consultants and business analysts	24.70	23.72	96.0
Senior police officers	30.96	30.08	97.2
Social workers	20.39	20.21	99.1
Bar staff	8.33	9.03	108.4
All occupations	**19.70**	**17.35**	**88.1**
Average gross *weekly* pay (incl. overtime)	781.40	649.00	83.1
Average *weekly* hours worked (incl. overtime)	39.70	37.50	
Average *weekly* overtime	1.10	0.50	

Source: Earnings and Hours Worked, occupation by four-digit SOC: ASHE Table 14.6a (ONS, October 2021)

A narrowing gap?

The chart shows how the gender wage gap (based on *median* gross hourly earnings for full-time employees, excluding overtime) has narrowed in recent years from 17.4 per cent in 1997 to 7.9 per cent in 2021.

The IFS research of 2016 suggests that an important reason for an overall decline in the gender gap is the increasing proportion of women who are now educated to 'A' level and degree standard. However, the same study also found that there had been little change in the gender pay gap for the highly educated. One reason identified for this is that, despite equal pay legislation, women are still discriminated against when it comes to promotion, especially to senior positions.

The chart also shows the gender gap by the highest and lowest earners. The gap for the top 10 per cent of earners is higher and has narrowed more slowly, from 20.7 per cent in 1997 to 16.1 per cent in 2021. For the lowest 10 per cent of earners, the gap has fallen from 13.6 per cent in 1997 to 7.9 per cent in 2021. This is likely to be related to the introduction of minimum wage legislation (see Box 7.4) which will have disproportionately benefited women who are more likely to work in lower-paid occupations.

UK government initiatives

The UK government established the Davies review in 2010[2] to examine the under-representation of women on the boards of FTSE 100 and 350 companies. In 2011, only 12.5 per cent of FTSE 100 board members were female. The first Davies report recommended a number of initiatives for businesses to improve this situation and set a target of 25 per cent by 2015. This figure was met (26.1 per cent in 2015) and Lord Davies set a new target of 33 per cent by 2020. Despite COVID-19, this target was surpassed and a government update in February 2022 showed that almost 40 per cent of board members were female. This placed the UK in second place in the international rankings for female representation on boards.

The government has also introduced a number of other initiatives to try to close the gender pay gap including:

- Shared parental leave. This was introduced in April 2015 and enables parents to share up to 50 weeks of leave, of which 37 is paid. At present, the take-up of the scheme has been low – approximately 5 per cent.
- Requirement on larger firms to publish pay data. Companies employing more than 250 people will have to publish information about their employee pay as it was on 5 April 2017. The information must include the mean gender gap for basic pay as well as bonuses. It also has to include the proportion of men and women in each quarter of the pay structure. Firms had until April 2018 to produce the data or they would be contacted by the Equalities and Human Rights Commission.
- Women in Finance Charter. This is a pledge for 'a more balanced and fair' financial services sector. In November 2017, HM Treasury announced that 162 firms had signed the Women in Finance Charter and that over 85 per cent of signatories had committed to have at least 30 per cent women in senior roles by 2021 and over 25 per cent had committed to a 50/50 gender split in senior roles by 2021.

It will be interesting to see what impact these policies have on the gender pay gap in the future.

1. *If employers were forced to give genuinely equal pay for equal work, how would this affect the employment of women and men? What would determine the magnitude of these effects?*
2. *How could family policy ensure that parents are able to work, while reducing pay differentials?*
3. *What measures could a government introduce to increase the number of women getting higher-paid jobs?*

Download data on the Gender Pay Gap from the labour market earnings database on Eurostat. Using the latest available data, create a chart showing the variation in the size of the gender pay gap across a selection of EU member countries along with the gap for the UK.

[2] *Improving the Gender Balance on British Boards*, Davies Review Five Year Summary (October 2015).

Table 7.3 Sources of British household wealth: April 2018–March 2020

	Total, £ billion	% of total	Median, £
Net property wealth	5458	35.9	112 000
Physical wealth	1385	9.1	38 000
Net financial wealth	1933	12.7	8000
Private pension wealth	6 445	42.3	111 700
Total wealth	**15 211**	**100.0**	**302 500**

Source: Household total wealth in Great Britain: April 2018 to March 2020 (ONS, January 2022)

Since the late 1970s, wage differentials have widened. Part of the explanation lies in a shift in the demand for labour. Many firms have adopted new techniques that require a more highly educated workforce. Wage rates in some of these skilled occupations have increased substantially.

At the same time there has been a decline in the number of unskilled jobs in industry, and along with it, a decline in the power of unions to represent such people. Where low-skilled jobs remain, there is intense pressure on employers to reduce wage costs if they are competing with companies based in developing countries, where wage rates are much lower.

As prospects for the unskilled decline in industry, so people with few qualifications increasingly compete for low-paid service-sector jobs (e.g. in supermarkets, fast-food outlets and call centres). The growth in people seeking part-time work has also kept wage rates down in this sector.

The distribution of wealth

In 2020/21 the median level of disposable income in the UK was £31 400, while the median level of household wealth over the period 2018–2020 was £302 500. Therefore, median wealth was 9.6 times larger than median annual disposable income.

Table 7.3 shows the composition of household wealth in Great Britain for 2018–2020. The largest by value is *private pension wealth* (42.3 per cent of total wealth). This is the accrued value of private pensions, such as occupational pension or personal pension. The second largest is *net property wealth* (35.9 per cent of total wealth). This is the value of a household's main residence plus any other property or land they may own. Next largest is *physical wealth* (12.7 per cent). This includes vehicles and contents and valuables within property. The smallest component is *net financial wealth* (9.1 per cent) which is the value of financial assets (savings, etc.) less financial liabilities (debts).

KI 24 p 185

Figure 7.12 shows how the inequality of wealth is far greater than inequality of income in Britain. It plots Gini coefficients for three different measures of income in 2020/21, reflecting different stages of redistribution, and also for total wealth and its components for the period from 2018 to 2020.

The Gini coefficient for total wealth is estimated at 0.62 compared to 0.486 for original income (before government intervention) and 0.344 for disposable income (after receipt of cash benefits and payment of direct taxes). The Gini coefficients show net financial wealth to be the most unequally distributed component of wealth (0.89) and physical wealth the least (0.47).

Causes of wealth inequality

The four major causes of inequality in the distribution of wealth are as follows:

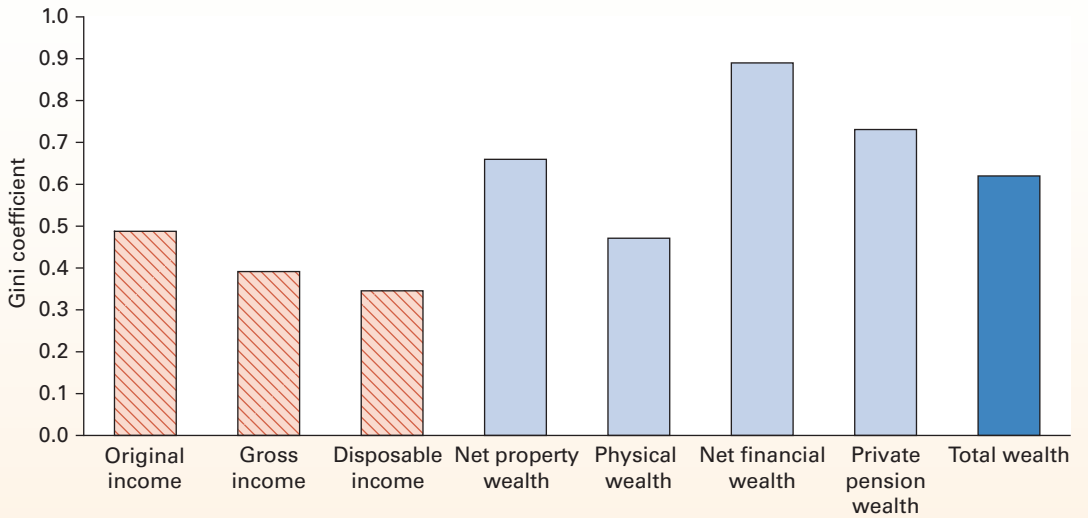
- *Social immobility and inheritance.* **Social mobility** relates to the ability of people from disadvantaged backgrounds to move up in the world. Hence, social immobility can perpetuate inequalities, such as those in wealth, from one generation to another.
- *Income inequality.* People with higher incomes can save more.
- *Different propensities to save.* People who save a larger proportion of their income will build up a bigger stock of wealth.

KI 24 p 185

Definition

Social mobility The ease with which people can move within or between social strata, such as those defined by wealth, income, class, occupation, etc.

Figure 7.12 UK Gini coefficients for income (2020/21) and wealth (2018–2020)



Source: Based on data in *The effects of taxes and benefits on household income, UK, 2020/21*, Reference Table 9 (ONS, March 2022) and *Household total wealth in Great Britain: April 2018 to March 2020* Table 2.3 (ONS, January 2022)

- *Entrepreneurial and investment talent/luck*. Some people are successful in investing their wealth and making it grow rapidly.

Pause for thought

What measures could be used to capture the degree of social mobility?

Even though wealth is still highly concentrated, there was a significant reduction in inequality of wealth up to the early 1990s. From 1971 to 1991 the Gini coefficient of wealth fell a full 16 percentage points from 0.80 to 0.64. A major reason for this was the increased taxation of inherited wealth. Since then, however, the reduction in inequality has largely stalled.

Between 2006 and 2020, the Gini coefficient of wealth fluctuated between 0.61 and 0.63. During this period the small reduction in the inequality of pension wealth (0.77 to 0.73) was offset by a more unequal distribution of net property wealth (0.62 to 0.66).

Causes of inequality

We turn now to identify the major causes of inequality. The problem has many dimensions and there are many factors that determine the pattern and depth of inequality. It is thus wrong to try to look for a single cause, or even the major one. The following are possible determinants of inequality:

- *Inequality of wealth*. People with wealth are able to obtain an income other than from their own labour (e.g. from rent or dividends on shares).
- *Differences in ability*. People differ in intelligence, strength, etc. Some of these differences are innate and some are acquired through the process of 'socialisation' – education, home environment, etc.
- *Differences in attitude*. Some people are adventurous, willing to take risks, willing to move for better jobs, keen to push themselves forward. Others are much more cautious. KI 15 p 80
- *Differences in qualifications*. These are reflections of a number of things: ability, attitudes towards study, access to good education, income of parents, etc.
- *Differences in hours worked*. Some people do a full-time job plus overtime, or a second job; others work only part time.
- *Differences in job utility/disutility*. Other things being equal, unpleasant or dangerous jobs will need to pay higher wages.
- *Differences in power*. Monopoly power in the supply of factors or goods, and monopsony power in the demand for factors, is unequally distributed in the economy.
- *Differences in the demand for goods*. Workers employed in expanding industries will tend to have

BOX 7.4 MINIMUM WAGE LEGISLATION

A way of helping the poor?

The UK Labour government introduced a statutory UK minimum wage in April 1999. For the year from April 2022, the rates were £9.50 per hour for those aged 23 and over – the so-called National Living Wage (NLW); £9.18 for those aged 21 and 22; £6.83 for those aged 18 to 20; and £4.81 for under-18s. The government has set a target for the NLW to reach two-thirds of median earnings by 2024.

As of March 2022, 21 of the 27 EU member countries had a national minimum wage. These rates varied considerably, but three broad groupings can be identified: one where minimum wages were lower than €1000 a month, which comprised mainly eastern EU member states (Bulgaria, Romania, Lithuania, the Czech Republic, Hungary, Latvia, Slovakia, Estonia, Croatia and Poland) plus Greece, Malta and Portugal. A second small group with minimum wages between €1000 and €1500 per month (Slovenia and Spain) and a third with minimum wages above €1500 per month (France, Germany, Belgium, the Netherlands, Ireland and Luxembourg).

Luxembourg has the highest minimum wage in the EU at €2257, which is almost 6.5 times higher than the lowest, Bulgaria's, which is €332. However, if these rates are adjusted to take into account price levels in each country, Luxembourg's minimum wage becomes just 2.8 times higher than that in Bulgaria.

Other countries also have minimum wages in place, including Japan, where the rate was ¥930/hour (£6.14/hour) in 2022, having increased from ¥771.33/hour in 2010. In the USA, there is a federal minimum wage of $7.25, which has remained unchanged since 2009, but 30 of the 50 states set higher rates. In 2022, Washington DC's rate was the highest at $15.20, followed by California at $15.00.

As with the USA, minimum wages in China vary across the country, often reflecting different stages of development, a rural or urban location, prices and the cost of living. Minimum wages have been increasing, but they still remain comparatively low. Beijing has the highest minimum wage at 25.3 yuan per hour (£3.05), while Shandong's wage is between 17 and 21 yuan per hour (£2.05–£2.53) and Liaoning's is between 14.3 and 19.2 yuan per hour (£1.72– £2.31).

Assessing the arguments

The principal argument *against* imposing a national minimum wage concerns its impact on employment. If you raise wage rates above the equilibrium level, there will be surplus labour: i.e. unemployment (see Figure 7.6 on page 169). However, the impact of a national minimum wage on employment is not so simple.

In the case of a firm operating in *competitive* labour and goods markets, the demand for low-skilled workers is relatively wage sensitive. Any rise in wage rates, and hence prices, by this firm alone would lead to a large fall in sales and hence in employment. But given that *all* firms face the minimum wage, individual employers are more able to pass on higher wages in higher prices, knowing that their competitors are doing the same.

When employers have a degree of monopsony power, however, it is not even certain that they would want to reduce employment. Remember what we argued in Box 7.2 (on page 170) when we were examining the effects of unions driving up wages. The argument is the same with a minimum wage. The minimum wage can be as high as W_2 and the firm will still want to employ as many workers as at W_1. The point is that the firm can no longer drive down the wage rate by employing fewer workers, so the incentive to cut its workforce has been removed.

In the long run, the effect on unemployment will depend on the extent to which the higher wages are compensated by higher *labour productivity*.

a higher marginal revenue product because their output has a higher market value.

- *Differences in household composition*. The greater the number of dependants relative to income earners, the poorer the average household member will be (other things being equal). Similarly, the greater the number of retired people relative to economically active individuals, the poorer the household will tend to be.
- *Discrimination* by race/ethnicity, sex/gender, age, social background, etc.
- *Degree of government support*. The greater the support for the poor, the less will be the level of inequality in the economy. TC 5 p 17
- *Unemployment*. When unemployment levels are high, this is one of the major causes of poverty.

Pause for thought

Which of these causes of inequality are reflected in differences in the marginal revenue product of labour?

CASE STUDIES & APPLICATIONS

Evidence from the USA and other countries suggests that modest increases in the minimum wage have had a neutral effect upon employment. Similarly in the UK, there is little evidence to suggest that from 1999 employers responded by employing fewer workers. In fact, until 2008, and again from 2013 to 2019, unemployment rates fell, due to the buoyant economy and increasing labour market flexibility.

Whether there would continue to be little effect if the minimum wage were to rise substantially is another matter. The issue, then, seems to be how *high* can the minimum wage be set before unemployment begins to rise.

Impact of the minimum wage in the UK

The Low Pay Commission estimates that in April 2021, approximately 1.42 million workers (5.4 per cent of workers) were in receipt of the NLW and 275 000 (13.6 per cent) of the various under-23 rates.

Evidence suggests that that there is also a *ripple effect* on workers earning just above the minimum wage rates. Employers find that they have to pay a sufficient premium above the NLW to attract more skilled workers and therefore a rise in the NLW pushes up these rates too.

There has been criticism of the UK government for using the term 'National Living Wage'. The term 'living wage' or 'Real Living Wage (RLW)' has been used for several years by the Living Wage Foundation. It is independently calculated each year by the Resolution Foundation 'based on what employees and their families need to live'. It is paid by around 9000 UK employers and in 2021/22 was estimated at £11.05 in London and £9.90 in other parts of the UK for all workers over 18. A large increase in the RLW was expected for 2022/23 to cope with the large rise in energy and basic food prices.

The fact that living costs, and particularly housing costs, vary so substantially across the country has led to suggestions that *local* rather than national minimum wages rates should be set to ensure both a fairer and a more efficient allocation of resources.

Of course a weakness of using a minimum wage as a means of relieving poverty is that it only affects the employed; yet one of the main causes of poverty is unemployment. This, in part, explains the focus on getting adults into work as the major plank of poverty-relief policy in the UK. It is apparent that a minimum wage rate cannot be the sole answer to poverty and must be considered in conjunction with benefits.

1. *If an increase in wage rates for low-paid workers leads to their being more motivated, how would this affect the marginal revenue product and the demand for such workers? What implications does your answer have for the effect on employment in such cases? (See page 171 on the efficiency wage hypothesis.)*
2. *If a rise in the minimum wage encourages employers to substitute machines for workers, will this necessarily lead to higher long-term unemployment in (a) that industry and (b) the economy in general?*

By visiting the GOV.UK website, collect data on UK National Minimum Wage and National Living Wage rates over time. Chart the data according to age group. In discussing your findings, think about any additional data you would need to analyse how the spending power of these groups will have changed over time.

Recap

1. Inequality can be examined by looking at the size distribution of income, the functional distribution of income, the distribution of income by recipient and the distribution of wealth.
2. The distribution of income and wealth can be illustrated by means of a Lorenz curve. The greater the inequality, the more bowed the curve will be towards the bottom right-hand corner.
3. The distribution can also be measured by a Gini coefficient. This will give a figure between 0 (total equality) and 1 (total inequality).
4. Wages and salaries constitute by far the largest source of income, and thus inequality can be explained mainly in terms of differences in wages and salaries. Nevertheless, state benefits are an important moderating influence on inequality and constitute the largest source of income for the poorest 20 per cent of households. Investment earnings are only a minor determinant of income except for some of the very richest households.
5. Inequality in wages and salaries largely reflects differences in the productivity of workers, consumer demand, power and discrimination. Other determinants of income inequality include differences in household composition, inequality of wealth, unemployment and the level of government benefits.
6. The distribution of wealth is less equal than the distribution of income.

BOX 7.5 INEQUALITY AND ECONOMIC GROWTH

CASE STUDIES & APPLICATIONS

Macroeconomic implications of income inequality

What is the relationship between inequality in a country and its rate of economic growth? The traditional answer is that there is a trade off between the two. Increasing the rewards to those who are more productive or who invest encourages a growth in productivity and capital investment, which, in turn, leads to faster economic growth. Redistribution from the rich to the poor, by contrast, is argued to reduce incentives by reducing the rewards from harder work, education, training and investment. Risk taking, it is claimed, is discouraged.

Various research papers, however, challenge the traditional view. Work by Ostry, Berg, Charalambos and Tsangarides[1] for the IMF suggests that 'lower net inequality is robustly correlated with faster and more durable growth, for a given level of redistribution'.

Research from the OECD[2] in the context of an increase in the OECD average Gini coefficient from 0.29 in the mid 1980s to 0.32 in 2011/12 suggests that a 3 percentage point rise in the Gini Coefficient 'would drag down growth by 0.35 percentage points per year for 25 years: a cumulative loss in GDP at the end of the period of 8.5 per cent'.

Social mobility and social capital

But why should a rise in inequality lead to lower economic growth?

One argument is that inequality in income and wealth affects opportunity and *social mobility* (page 180). Disadvantaged groups may not have the opportunities, particularly in education and health, which would enhance their 'human capital': i.e. their knowledge, skills, competencies and other attributes. For example, people from poorer backgrounds on average leave school or college earlier and with lower qualifications. (Human capital is discussed further in Box 10.4.)

By restricting the development of human capital, such inequalities limit the capacities of societies to produce goods and services and to generate and foster ideas, innovations and new technologies and ways of doing things. Importantly, the extent to which this restricts the growth of an economy's productive potential impacts, not only on the current size of the national cake, but on the size of tomorrow's cake too.

It is also argued by some that inequality and social immobility erode ***social capital***, which adversely impacts on economic growth. Social capital is the social connections, whether through relationships with family and friends, local communities or wider society, that help to bind societies together. In work published by the IMF in 2017, it is argued that 'excessive levels of inequality can erode social cohesion, lead to political polarisation, and ultimately lower economic growth'.[3]

But if greater inequality generally results in lower economic growth, will a redistribution from rich to poor necessarily result in faster economic growth? According to the OECD[4]:

> Anti-poverty programmes will not be enough. Not only cash transfers but also increasing access to public services, such as high-quality education, training and health care, constitute long-term social investment to create greater equality of opportunities in the long run.

Thus redistribution policies need to be well designed and implemented and focus on raising incomes of the poor through increased opportunities to increase their productivity. Simple transfers from rich to poor via the tax and benefits system may, in fact, undermine economic growth. According to Ostry, Berg, Charalambos and Tsangarides[5]:

> That equality seems to drive higher and more sustainable growth does not in itself support efforts to redistribute. In particular, inequality may impede growth at least in part because it calls forth efforts to redistribute that themselves undercut growth. In such a situation, even if inequality is bad for growth, taxes and transfers may be precisely the wrong remedy.

Identify policy measures that would both reduce inequality and increase income growth.

Prepare a short briefing paper for a government minister, making the case for policy measures that could both reduce inequality and increase income growth.

Definition

Social capital (OECD definition) Networks, together with shared norms, values and understandings, that facilitate co-operation within or among groups.

---

[1] Jonathan D. Ostry, Andrew Berg, and Charalambos G. Tsangarides, 'Redistribution, Inequality, and Growth', *IMF Staff Discussion Note* (February 2014).

[2] 'Does income inequality hurt economic growth?', *Focus on Inequality and Growth*, OECD (December 2014).

[3] 'Tackling Inequality, October 2017', *Fiscal Monitor*, IMF (October 2017).

[4] Op. cit.

[5] Op. cit.

7.4 THE REDISTRIBUTION OF INCOME

How can income be redistributed from rich to poor? What will be the effects of doing so?

TC 7 p 26

In this section we will look at policies to redistribute incomes more equally, and in particular we will focus on the use of government social security benefits and taxation. In doing so, we will be focusing on the economic and social goal of 'equity' and so the distributive role for government identified in Chapter 1 (see page 26).

KEY IDEA 24

Equity **is where income, wealth and access to health, education and other services are distributed in a way that is considered to be fair or just. Note that an equitable distribution is not the same as a totally equal distribution and that different people have different views on what is equitable.**

Taxation

If taxes are to be used as a means of achieving greater equality, the rich must be taxed proportionately more than the poor. The degree of redistribution will depend on the degree of 'progressiveness' of the tax. In this context, taxes may be classified as follows:

- **Progressive tax**. As people's income (Y) rises, the percentage of their income paid in the tax (T) rises. In other words, the average rate of tax (T/Y) rises. Income taxes are progressive (but much less progressive in the UK than they used to be).
- **Regressive tax**. As people's income rises, the percentage of their income paid in the tax falls: T/Y falls.
- **Proportional tax**. As people's income rises, the percentage of their income paid in the tax stays the same: T/Y is constant.

Therefore, progressiveness is defined in terms of what happens to the *average* rate of tax as income rises. The more progressive a tax, the more it will redistribute incomes away from the rich. Regressive taxes will have the opposite effect, since they tax the rich proportionately less than the poor.

Definitions

Progressive tax A tax whose average rate with respect to income rises as income rises.

Regressive tax A tax whose average rate with respect to income falls as income rises.

Proportional tax A tax whose average rate with respect to income stays the same as income rises.

Details of tax rates in the UK are given in Case Study 7.13 on the student website. This case study also examines how progressive or regressive the various types of tax are.

Problems with using taxes to redistribute incomes

How successfully can taxes redistribute income and at what economic cost?

Taxation takes away income. It can thus reduce the incomes of the rich. But no taxes, however progressive, can *increase* the incomes of the poor. This will require subsidies (i.e. benefits).

But what about tax cuts? Can bigger tax cuts not be given to the poor? This is possible only if the poor are already paying taxes in the first place. Take the two cases of income tax and taxes on goods and services.

- *Income tax*. If the government cuts income tax, then anyone currently paying it will benefit. A cut in tax *rates* will give proportionately more to the rich, since they have a larger proportion of taxable income relative to total income. An increase in personal *allowances*, on the other hand, will give the same *absolute* amount to everyone above the new tax threshold. This will therefore represent a smaller proportionate gain to the rich. In either case, however, there will be no gain at all to those people below the tax threshold. They paid no income tax in the first place and therefore gain nothing at all from income tax cuts.
- *Taxes on goods and services*. Since taxes such as VAT and excise duties on alcoholic drinks, tobacco, petrol and gambling are generally regressive, any cut in their rate will benefit the poor proportionately more than the rich. A more dramatic effect would be obtained by cutting the rate most on those goods consumed relatively more by the poor (e.g. on domestic fuel).

The government may not wish to cut the overall level of taxation, given its expenditure commitments. In this case, it can switch the burden of taxes from regressive to progressive taxes, if it wishes to benefit the very poor.

Taxation and incentives

TC 6 p 24

Tax changes affect incomes and/or prices. This, in turn, will change the incentives to consume and produce. Higher income taxes, for example, could discourage people from working. Higher taxes on certain products could discourage their consumption and hence production.

The whole question of taxation, choices and incentives can be highly charged politically. For instance, many countries, including the UK, have attempted since the 1980s to reduce both the basic and higher rates of income tax. While this has implications for the distribution of income, the primary economic objective behind the tax changes has been to increase the incentive to work. According to the economic right, although high and progressive income taxes can lead to a more equal distribution of income, they are likely to result in a smaller national output. Alternatively, by cutting taxes there will be a bigger national output but less equally divided. If many on the political left are correct, however, by raising income taxes we can have both a more equal society and a *bigger* national output: there is no trade-off.

The key to analysing these arguments is to distinguish between the *income effect* and the *substitution effect* of raising taxes. Raising income tax does two things.

- It reduces disposable incomes. People therefore are encouraged to work *more* in an attempt to maintain their consumption of goods and services. This is called the **income effect**. 'I have to work more to make up for the higher taxes', a person might say.
- It reduces the opportunity cost of leisure. Since higher income taxes reduce take-home pay, an extra hour taken in leisure now involves a smaller sacrifice in consumption. Thus people may substitute leisure for consumption, and work less. This is called the **substitution effect**. 'What is the point of doing overtime', another person might say, 'if so much of the overtime pay is going in taxes?'

TC 1 p 7
TC 6 p 24

The relative size of the income and substitution effects is likely to differ for different types of people. For example, the *income* effect is likely to dominate for those people with a substantial proportion of long-term commitments, such as those with families, with mortgages and other debts. They may feel forced to work *more* to maintain their disposable income. Clearly for such people, higher taxes are *not* a disincentive to work. The income effect is also likely to be relatively large for people on higher incomes, for whom an increase in tax rates represents a substantial cut in income.

The *substitution* effect is likely to dominate for those with few commitments: those whose families have left home, the single, and second income earners in families where that second income is not relied on for 'essential' consumption. A rise in tax rates for these people is likely to encourage them to work less.

TC 6 p 24

Although high income earners may work more when there is a tax *rise*, they may still be discouraged by a steeply progressive tax *structure*. If they have to pay very high marginal rates of tax, it may simply not be worth their while seeking promotion or working harder (see Case Study 7.15 on the student website).

One final point should be stressed. For many people there is no choice in the amount they work. The job they do dictates the number of hours worked, irrespective of changes in taxation.

Pause for thought

How will tax cuts affect the willingness of women with employed partners to return to paid work after having brought up a family? (Clue: think about the size of the income and substitution effects.)

Benefits

Benefits can be either cash benefits or benefits in kind.

Cash benefits

Means-tested benefits. **Means-tested benefits** are available only to those whose income (and savings in some instances) fall below a certain level. To obtain such benefits, therefore, people must apply for them and declare their personal circumstances to the authorities.

The benefits could be given as grants or merely as loans. They could be provided as general income support or for the meeting of specific needs, such as rents, fuel bills and household items.

Universal benefits. **Universal benefits** are those that everyone is entitled to, irrespective of their income, if they fall into a certain category or fulfil certain conditions. Examples include state pensions, and unemployment, sickness and invalidity benefits.

Figure 7.13 shows the expenditure on social cash benefits in selected countries since 1995. These include unemployment, sickness, invalidity, maternity, family, survivors' and housing benefits and state pensions. As you can see, benefits as a percentage of GDP (national income) vary from one country to another. For instance,

Definitions

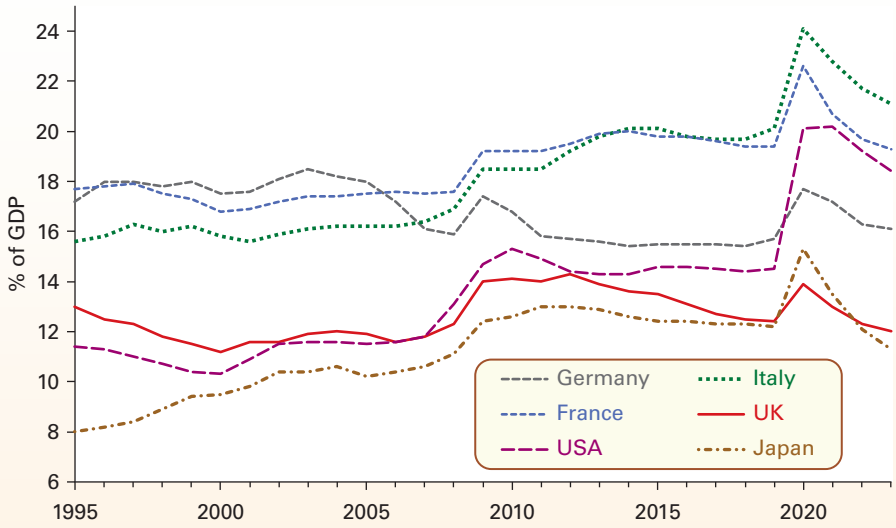
Income effect of a tax rise Tax increases reduce people's incomes and thus encourage people to work more.

Substitution effect of a tax rise Tax increases reduce the opportunity cost of leisure and thus encourage people to work less.

Means-tested benefits Benefits whose amount depends on the recipient's income or assets.

Universal benefits Benefits paid to everyone in a certain category irrespective of their income or assets.

Figure 7.13 Social cash benefits as a percentage of GDP



Note: Data from 2022 are based on forecasts

Source: Based on data in *AMECO Database* (European Commission)

the percentage shares in Japan, the USA and the UK are significantly below those in France and Italy. Note also the rise in benefits in 2020 as people were supported during lockdowns.

Differences between countries reflect a multitude of factors: some economic, but others not. Among the reasons are: differences in the coverage and levels of particular benefits, rates of unemployment, incomes and the age structure of the population. But cultural and historic traditions can matter too, such as views concerning the role of governments – a theme we shall pick up again in Chapter 8.

Benefits in kind

Individuals receive other forms of benefit from the state, not as direct monetary payments, but in the form of the provision of free or subsidised goods or services. These are known as **benefits in kind**. The two largest items in most countries are health care and education. They are very differently distributed, however. This difference can largely be explained on age grounds. Old people use a large proportion of health services, but virtually no education services.

Benefits in kind tend to be consumed roughly equally by the different income groups. Nevertheless they do have some equalising effect, since they represent a much larger proportion of poor people's income than rich people's. They have a far smaller redistributive effect, however, than cash benefits.

The system of benefits in the UK and their redistributive effects are examined in Case Study 7.16 on the student website.

Definition

Benefits in kind Goods or services which the state provides directly to the recipient at no charge or at a subsidised price. Alternatively, the state can subsidise the private sector to provide them.

Benefits and the redistribution of income

It might seem that means-tested benefits are a much more efficient system of redistributing income from the rich to the poor: the money is directed to those most in need. With universal benefits, by contrast, many people may receive them who have little need for them. Do families with very high incomes need child benefit? Would it not be better for the government to redirect the money to those who are genuinely in need? Indeed, in the UK, from 2012, families with at least one higher-rate taxpayer were no longer eligible for child benefit.

KI 20 p 127

KI 24 p 185

There are, however, serious problems in attempting to redistribute incomes by the use of means-tested benefits:

- Not everyone entitled to means-tested benefits applies for them, whether from ignorance of what is available, from the complexities of claiming or from reluctance to reveal personal circumstances. Thus some of the poorest families may receive no support.
- The levels of income above which people become ineligible for benefits may be set too low. Even if they were raised, there will always be some people just above these levels who will still find difficulties.
- Means tests based purely on *income* (or even universal benefits based on broad categories)

BOX 7.6 UK TAX CREDITS

An escape from the poverty trap?

Although minimum and 'living' wages help the problem of low pay in the UK, there are other policies that help those on low incomes. In 2003, Working Tax Credit (WTC) and Child Tax Credit (CTC) replaced Working Families Tax Credit.

WTC is a cash benefit supporting people on low incomes, but it also incentivises people to work more hours to meet the eligibility threshold. To be eligible for the basic amount of WTC (£2070 per year in 2022/23), claimants must be over 25 and work at least 30 hours per week (16 hours per week for single parents, those over 60 or the disabled). Additional payments are available depending on individual circumstances, which include support for 70 per cent of eligible childcare costs. The benefit, therefore, also helps parents who might otherwise struggle to work.

CTC provides support to families with children. Unlike WTC, eligibility is not dependent on someone working in the family and is paid in addition to WTC. In 2022/23, for children born after 7 April 2017, families receive up to £2935 per child up to a maximum payment of £5870. Payments of both tax credits fall once income exceeds a threshold level, helping to target the benefit.

During COVID-19, the basic element of WTC temporarily rose to support households, with claimants receiving additional payments depending on their circumstances. In April 2011, over 6 million families were claiming CTC and/or WTC. This number has fallen steadily, dropping to under 5 million with the introduction of the replacement system, Universal Credit (UC), and falling further when its roll-out was digitised, with new claimants moved from WTC/CTC to UC. As of April 2021, 1.9 million families were claiming CTC and/or WTC – a fall of 427 000 from 2020.[1]

Universal Credit

Universal Credit (UC) aimed to 'simplify the system to make work pay and combat worklessness and poverty'.[2] This means-tested benefit replaced WTC/CTC, plus Housing Benefit, Income Support, income-based Jobseeker's Allowance and income-related Employment and Support Allowance. It provides a single payment to support people in work, out of work, those with children and those with disabilities.

UC initially sets a claimant's maximum entitlement. This includes a standard allowance (dependent on age/circumstances) and additional allowances for children, housing costs, childcare costs and disabilities. In 2022/23[3] the monthly standard allowance for a person aged under 25 was £265.31; for a couple both aged over 25, it was £525.72. The additional allowance for a child born after 6 April 2017 was £244.58 and people could claim up to £646.35 per month for childcare costs. The next stage is calculating a *work allowance* – the amount that can be earned before payments are deducted. In 2022/23 this was £344 if the UC claim included housing support and £573 if it did not.

As of 13 January 2022, 5.6 million people were claiming UC – a marked increase from the 3 million people claiming it in March 2020. Much of the increase was due to COVID-19, with incomes falling and many individuals losing their jobs or being furloughed. Between 15 October 2021 and 13 January 2022, an average of 29 000 new claims were made per week – significantly lower than the 300 000 weekly average of new claims made in the four weeks to 9 April 2020. The average number of claims per week was falling and it was likely that this trend would continue as employment rose with the ongoing recovery.

Benefit systems and incentives

In designing benefit systems, it is important to consider their impact on work incentives. As people's incomes rise, their entitlement to benefits falls and this can discourage work. As the text at the bottom of this page explains, this is called the *poverty trap*.

[1] *Child and Working Tax Credits Statistics*, HM Revenue & Customs (23 June 2021).

[2] *Universal Credit: Welfare that works*, DWP (November 2010).

[3] *Benefit and Pension Rates 2022 to 2023*, DWP (11 January 2022).

ignore the very special needs of many poor people. A person earning £150 a week living in a small, well-appointed flat with a low rent will have less need of assistance than another person who also earns £150 per week but lives in a cold, draughty and damp house with large bills to meet. If means tests are to be really fair, then *all* of a person's circumstances need to be taken into account.

KI 24 p 185

The tax/benefit system and the problem of disincentives: the 'poverty trap'

When means-tested benefits are combined with a progressive income tax system, there can be a serious problem of *disincentives*. As poor people earn more money, so not only do they start paying income taxes and national insurance, but also they begin losing means-tested benefits.

TC 6 p 24

Theoretically, it is possible to have a marginal tax-plus-lost-benefit rate in excess of 100 per cent. In other words, for every extra £1 earned, taxes and lost benefits add up to more than £1. High marginal tax-plus-lost-benefit rates obviously act as a serious disincentive. What is the point of getting a job or trying to earn more money if you end up earning little more or actually losing money?

This situation is known as the **poverty trap**. People are trapped on low incomes with no realistic means of bettering their position.

Definition

Poverty trap Where poor people are discouraged from working or getting a better job because any extra income they earn will be largely taken away in taxes and lost benefits.

CASE STUDIES & APPLICATIONS

Two key issues affect these incentives. The first is the income threshold at which entitlement begins to fall; the second is the rate at which it declines – the taper rate. A 40 per cent taper rate means that for every pound earned above the threshold, entitlement to the benefit falls by 40p.

A higher threshold when tapering begins and a lower taper rate increases work incentives for those on low incomes, as they lose fewer benefits by working or accepting a promotion/new job. However, it also means that families on higher incomes become eligible to receive payments, affecting their work incentives. A lower taper rate also increases the cost to government. Tax credits thus illustrate the problem of providing support to poor people which is affordable for taxpayers without creating disincentives to work. The more gently the support tapers off (and hence the less the disincentive to earn extra money), the costlier it is to finance and the higher the tax rates that are needed elsewhere.

Another key issue within any benefits system is its complexity. A complex system can reduce take-up of benefits, reduces work incentives if recipients are unclear about the impact of their decisions and is more administratively costly and prone to errors. For example, WTC and CTC both have a taper rate of 41 per cent, but the thresholds at which entitlement declines vary. In 2022/23, this is £6770 for WTC and £17 005 for CTC.

Universal credit aimed to address some of these concerns:

- A single taper rate reduces the impact of the poverty trap and increases work incentives. The UC taper rate was reduced from 63 per cent to 55 per cent in the Autumn 2021 Budget.
- Removing the requirement to work a set number of hours eradicates a distortion in the system.
- Bringing together in-work and out-of-work benefits reduces the risks and transactions costs for those moving from unemployment into work.
- Simplifying the claims process, with only one application to DWP needed, rather than multiple applications to different departments, reduces administrative costs.

Despite the benefits, the roll-out faced serious technical issues and delays. All 8 million claimants on the previous system were due to be moved to UC by October 2017, but only 540 000 were moved. There has been a steady increase in the numbers claiming UC since, but with further delays, a new target was set for September 2024.[4]

Other issues include delays between applying and receiving a payment, creating financial issues for applicants; monthly payments in arrears, which can cause issues if recipients are paid weekly; annual increases not keeping pace with rising costs, such as energy and food; and conditions such as work search activities/regular interviews which if not met, can reduce payments. A House of Lords Economic Affairs Committee inquiry called for UC to be reformed to take account of these problems.[5]

Economists sometimes refer to an 'unemployment trap'. People are discouraged from taking work in the first place. Explain how such a 'trap' arises. Has the introduction of Universal Credit in the UK created an unemployment trap? What are the best ways of eliminating, or at least reducing, the unemployment trap?

Conduct desktop research to investigate how COVID-19 affected the roll-out and take-up of Universal Credit. Consider whether issues that arose had to do with the design of the system, the incentive mechanisms involved or the impact of the pandemic.

[4] *"Universal Credit roll out delayed again – to 2024'*, BBC News (3 February 2020).

[5] *Universal Credit isn't working: proposals for reform*, House of Lords Economic Affairs Committee (July 2020).

The problem of the poverty trap would be overcome by switching to a system of universal benefits unrelated to income. For example, *everyone* could receive a flat payment from the state fixed at a sufficiently high level to cover their basic needs. There would still be *some* disincentive, but this would be confined to an income effect: people would not have the same need to work if the state provided a basic income. But there would no longer be the disincentive to work caused by a resulting *loss* of benefits (a substitution effect).

The big drawback with universal benefits, however, is their cost. If they were given to everyone and were large enough to help the poor, their cost would be enormous. Thus although the benefits themselves would not create much disincentive effect, the necessary taxation to fund them probably would.

There is no ideal solution to this conundrum. On the one hand, the more narrowly benefits are targeted on the poor, the greater the problem of the poverty trap. On the other hand, the more widely they are spread, the greater the cost of providing any given level of support to individuals.

A compromise proposal is that of a **negative income tax**, whereby below a certain level of income people pay a 'negative tax': in other words, they receive a benefit. The lower the income, the greater

Definition

Negative income tax A combined system of taxes and benefits. Below a certain level of income people pay a negative tax: in other words they receive benefits. As people earn more, they gradually lose their benefits until beyond a certain level they begin paying taxes – the tax becomes positive.

this benefit. This is examined in Case Study 7.17 on the student website. Box 7.6 examines the use of tax credits and Universal Credit – forms of negative income tax – in the UK.

Pause for thought

Does the targeting of benefits to those in greatest need necessarily increase the poverty trap?

Recap

1. Taxes can be categorised as progressive, regressive or proportional. Progressive taxes have the effect of reducing inequality. The more steeply progressive they are, the bigger the reduction in inequality.
2. Taxes on their own cannot increase the incomes of the poor. Cutting taxes, however, can help the poor if the cuts are carefully targeted.
3. Using taxes to redistribute incomes can cause disincentives. Raising taxes has two effects on the amount people wish to work. On the one hand, people will be encouraged to work more in order to maintain their incomes. This is the income effect. On the other hand, they will be encouraged to substitute leisure for income (i.e. to work less), since an hour's leisure now costs less in forgone income. This is the substitution effect. The relative size of the income and substitution effects will depend on the nature of the tax change. The substitution effect is more likely to outweigh the income effect for those with few commitments, for people just above the tax threshold of the newly raised tax and in cases where the highest rates of tax are increased.
4. Benefits can be cash benefits or benefits in kind. They can be universal or means tested. Universal benefits include child benefit, state pensions, unemployment benefits, and sickness and invalidity benefits. Benefits in kind include health care, education and free school meals.
5. Means-tested benefits can be specifically targeted to those in need and are thus more 'cost-effective'. However, there can be serious problems with such benefits, including: limited take-up, time-consuming procedures for claimants, some relatively needy people falling just outside the qualifying limit and inadequate account taken of all relevant circumstances affecting a person's needs.
6. The poverty trap occurs when the combination of increased taxes and reduced benefits removes the incentive for poor people to earn more. The more steeply progressive this combined system is at low incomes, the bigger the disincentive effect.

QUESTIONS

1. If a firm faces a shortage of workers with very specific skills, it may decide to undertake the necessary training itself. If, on the other hand, it faces a shortage of unskilled workers, it may well offer a small wage increase in order to obtain the extra labour. In the first case it is responding to an increase in demand for labour by attempting to shift the supply curve. In the second case it is merely allowing a movement along the supply curve.

   Use a demand and supply diagram to illustrate each case. Given that elasticity of supply is different in each case, do you think that these are the best policies for the firm to follow?
2. For what types of reason does the marginal revenue product differ between workers in different jobs?
3. Why, do you think, are some of the lowest-paid jobs the most unpleasant?
4. The wage rate a firm has to pay and the output it can produce vary with the number of workers as shown in the table below (all figures are hourly). Assume that output sells at £2 per unit.

Number of workers	1	2	3	4	5	6	7	8
Wage rate (AC_L) (£)	10	11	12	13	14	15	16	17
Total output (TPP_L)	10	22	32	40	46	50	52	52

   a. Copy the table and add additional rows for TC_L, MC_L, TRP_L and MRP_L. Put the figures for MC_L and MRP_L in the spaces between the columns.
   b. How many workers will the firm employ in order to maximise profits?
   c. What will be its hourly wage bill at this level of employment?
   d. How much hourly revenue will it earn at this level of employment?
   e. Assuming that the firm faces other (fixed) costs of £30 per hour, how much hourly profit will it make?
   f. Assume that the workers now form a union and that the firm agrees to pay the negotiated wage rate to all employees. What is the maximum to which the hourly wage rate could rise without causing the firm to try to reduce employment below that in (b) above? (See the diagram in Box 7.2.)
   g. What would be the firm's hourly profit now?
5. The figures shown in the table at the top of the next page are for a monopsonist employer. Fill in the missing figures for columns (3) and (4). How many workers should the firm employ if it wishes to maximise profits?

Number of workers (1)	Wage rate (£) (2)	Total cost of labour (£) (3)	Marginal cost of labour (£) (4)	Marginal revenue product (£) (5)
1	100	100		
			110	230
2	105	210		
			120	240
3	110	330		
				240
4	115			
				230
5	120			
				210
6	125			
				190
7	130			
				170
8	135			
				150
9	140			
				130
10	145			

6. To what extent can trade unions be seen to be (a) an advantage, (b) a disadvantage to (i) workers in unions, (ii) employers bargaining with unions, (iii) non-union members in firms where there is collective bargaining, (iv) workers in non-unionised jobs?
7. Identify four groups of workers, two with very high wages and two with very low wages. Explain why they get the wages they do.
8. Why do the world's top footballers earn millions of pounds per year, while the top lacrosse players are paid less than the equivalent of £30 000?
9. Do any of the following contradict marginal productivity theory: (a) wage scales related to length of service (incremental scales); (b) nationally negotiated wage rates; (c) discrimination; (d) firms taking the lead from other firms in determining this year's pay increase?
10. For what reasons is the average gross weekly pay of women only around 83 per cent of that of men in the UK (see Box 7.3)?
11. Does the existence of overtime tend to increase or decrease inequality?
12. Distinguish between proportional, progressive and regressive taxation. Could a progressive tax have a constant marginal rate?
13. If a person earning £20 000 per year pays £2000 in a given tax and a person earning £40 000 per year pays £3200, is the tax progressive or regressive? Explain.
14. A proportional tax will leave the distribution of income unaffected. Why should this be so, given that a rich person will pay a larger absolute amount than a poor person?
15. a. What is meant by social mobility?
    b. What impact could an increase in social mobility have on the inequality of income and wealth?
    c. How could greater social mobility affect an economy's longer-term economic growth?
16. Under what circumstances would a rise in income tax act as (a) a disincentive and (b) an incentive to effort?
17. Who is likely to work harder as a result of a cut in income tax rates – a person on a high income or a person on a low income? Why? Would your answer be different if personal allowances were zero?
18. What tax changes (whether up or down) will both have a positive incentive effect and also redistribute incomes more equally?
19. What is meant by 'the poverty trap'? Would a system of universal benefits be the best solution to the problem of the poverty trap?
20. How would you go about deciding whether person A or person B gets more personal benefit from each of the following: (a) an electric fire; (b) a clothing allowance of £x; (c) draught-proofing materials; (c) child benefit? Do your answers help you in deciding how best to allocate benefits?

Pearson | MyLab | Economics

Log on to MyLab Economics and complete the studyplan exercises for this chapter to see how much you have learnt and where you need to revise most. Make sure you access all the supporting textbook resources, including the online workbook, newsblog, audio animations, guided solutions and ebook.

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

7.1 **Labour as a factor of production.** An examination of some of the ethical consequences of treating labour as an 'input' into production.

7.2 **Flexible labour markets.** This looks at the changes in the ways that firms organise their workforces.

7.3 **Telecommuters.** This case study looks at the growth in the number of people working from home.

7.4 **Life at the mill.** Monopsony in Victorian times as reported by Friedrich Engels.

(Continued)

7.5 **Poverty in the past.** Extreme poverty in Victorian England.

7.6 **The rise and decline of the labour movement.** A brief history of trade unions in the UK.

7.7 **How useful is marginal productivity theory?** How accurately does the theory describe employment decisions by firms?

7.8 **Profit-sharing.** An examination of the case for and against profit-sharing as a means of rewarding workers.

7.9 **Holidays: good for workers; bad for employers?** An examination of holiday entitlements in the USA and Europe and their effects on workers and business.

7.10 **How can we define poverty?** This examines different definitions of poverty and, in particular, distinguishes between absolute and relative measures of poverty.

7.11 **How to reverse the UK's increased inequality.** Recommendations of the Rowntree Foundation.

7.12 **Adam Smith's maxims of taxation.** This looks at the principles of a good tax system as identified by Adam Smith.

7.13 **Taxation in the UK.** This case study looks at the various types of tax in the UK. It gives the current tax rates and considers how progressive the system is.

7.14 **The poll tax.** This case charts the introduction of the infamous poll tax (or 'community charge') in the UK and its subsequent demise.

7.15 **The Laffer curve.** This curve suggests that raising tax rates beyond a certain level will reduce tax revenue. But will cutting tax rates increase revenue?

7.16 **The system of benefits in the UK.** A description of the various benefits used in the UK and their redistributive effects.

7.17 **Negative income tax and redistribution.** This case looks at the possible effects of introducing negative income taxes (tax credits).

7.18 **The squeezed middle.** What have been the effects on people on 'middle incomes' by successive governments to support poor families?

7.19 **Increase life expectancy.** This examines the likely future impact on individuals and economies of people living longer.

WEB APPENDICES

7.1 **Deriving the individual's supply curve of labour.** This uses indifference analysis (see Web Appendix 2.2) to explain the shape of an individual worker's supply curve of labour and explains why at high wage rates the curve may become backward bending.

7.2 **Economic rent and transfer earnings.** This appendix examines a way of classifying the earnings of a factor of production and shows how these earnings depend on the price elasticity of supply of the factor.

Chapter 8

Market failures and government policy

To what extent should we rely on markets to allocate resources? In recent years, governments throughout the world have tended to put more reliance on markets as the means of allocating resources. Policies of privatisation, deregulation, cutting government expenditure and taxes have been widely adopted by governments of all political persuasions.

Despite this increased reliance on markets, markets often fail. The economic actions of individuals, firms and governments often have adverse effects on other people and on the environment. Some people grow rich while others struggle to make ends meet. At the same time, many aspects of our lives seem to be dominated by the interests of big business.

While both the extent and nature of government involvement in economies varies from one country to another, all governments are still expected to play a major role. This involvement ranges from supporting and regulating businesses, to the passing of laws to protect the individual, to providing or enabling access to key services such as education, health care and law and order, to social protection in the form of pensions and social security, to the provision of infrastructure, such as a country's rail track or road network, and to measures to tackle global warming and environmental degradation.

In this chapter we identify the various ways in which the market fails to look after society's interests (Sections 8.1–8.4). Then we look at how the government can set about putting right these failings (Sections 8.5–8.7). Then we look at some of the shortcomings of governments, and ask: should we have more or less intervention? In the final section we turn to problems of the environment as a case study in market failure and government intervention.

After studying this chapter, you should be able to answer the following questions:

- What is the meaning of a 'socially efficient' allocation of resources and to what extent will a perfectly competitive market achieve social efficiency?
- For what reasons do real-world markets fail to achieve social efficiency?
- How can governments put right the failings of the market?
- How successful are they likely to be?
- How can economists explain environmental degradation?
- What policies can be pursued to achieve environmental sustainability?

8.1 SOCIAL EFFICIENCY

Is this something that the free market will achieve?

In order to decide the optimum amount of government intervention, it is first necessary to identify the various social goals that intervention is designed to meet. Two of the major objectives of government intervention identified by economists are **social efficiency** and **equity**.

Equity

KI 24 p 185

Most people would argue that the free market fails to lead to a *fair* distribution of resources, if it results in some people living in great affluence while others live in dire poverty. In Chapter 7 we considered how economists look to measure the degree of inequality in the distribution of income and wealth. Clearly what constitutes 'fairness' is a highly contentious issue: those on the political right generally have a quite different view from those on the political left. Nevertheless, most people would argue that the government does have some duty to redistribute incomes from the rich to the poor through the tax and benefit system, and perhaps to provide various forms of legal protection for the poor (such as a minimum wage rate).

In other words, government has a *redistributive* role to play (see page 26). Having looked at the causes of inequality and policies of redistribution in Chapter 7. we focus here on the second issue: that of social efficiency.

Social efficiency

TC 2 p 9

Remember how we defined 'rational' choices. A rational person will choose to do more or less of an activity if the gain from so doing exceeds any sacrifice involved. In other words, if the marginal benefit of an activity exceeds the marginal cost (i.e. the price), the rational person will do more of it; if marginal cost exceeds marginal benefit, the rational person will do less. But as people do more of an activity, its marginal benefit is likely to decline (the principle of diminishing marginal utility). When marginal benefit has fallen to equal marginal cost, it is rational to stop doing more; the optimum point level has been reached.

Similarly a 'rational' (i.e. profit-maximising) firm will produce more if marginal benefit (i.e. marginal revenue) exceeds marginal cost and produce less if marginal cost exceeds marginal revenue. Profit is maximised where $MR = MC$.

Thus the economist's rule for **rational economic behaviour** is that a person should expand or contract the level of any activity until its marginal benefit is equal to its marginal cost. At that point, the person will be acting efficiently in his or her own private interest. Only when $MB = MC$ can no further gain be made. This is known as a situation of **private efficiency**.

By analogy, *social* efficiency will be achieved where, for any activity, the marginal benefit to *society* (*MSB*) is equal to the marginal (opportunity) cost to *society* (*MSC*). If the marginal benefits to society – or 'marginal social benefits' (*MSB*) – of producing or consuming any given good or service exceed the marginal costs to society – or 'marginal social costs' (*MSC*) – then it is said to be socially efficient to produce or consume more. For example, if people's gains from having additional motorways exceed *all* the additional costs to society (both financial and non-financial), then it is socially efficient to construct more motorways.

KI 22 p 153

If, however, the marginal social costs of producing or consuming any good or service exceed the marginal social benefits, then it is socially efficient to produce or consume less.

It follows that if the marginal social benefits of any activity are equal to the marginal social costs, then the current level is the optimum. To summarise: to achieve social efficiency in the production or consumption of any good or service, the following should occur:

TC 2 p 9

$MSB > MSC \rightarrow$ produce or consume more

$MSC > MSB \rightarrow$ produce or consume less

$MSB = MSC \rightarrow$ keep production or consumption at its current level

Social efficiency is an example of 'allocative efficiency': in other words, the best allocation of resources between alternative uses.

Definitions

Social efficiency Production and consumption at the point where $MSB = MSC$.

Equity A fair distribution of resources.

Rational economic behaviour Doing more of those activities whose marginal benefit exceeds their marginal cost and doing less of those activities whose marginal cost exceeds their marginal benefit.

Private efficiency Where a person's marginal benefit from a given activity equals the marginal cost.

KEY IDEA 25

Allocative efficiency in any activity is achieved where any reallocation would lead to a decline in net benefit. It is achieved where marginal benefit equals marginal cost. Private efficiency is achieved where marginal private benefit equals marginal private cost ($MB = MC$). Social efficiency is achieved where marginal social benefit equals marginal social cost ($MSB = MSC$).

TC 12

The concept of allocative efficiency is another of our threshold concepts (no. 12). It is a threshold concept, because to understand it is to understand how to make the most of scarce resources: and scarcity is the core problem of economics for all of us.

In the real world, however, the market rarely leads to social efficiency: the marginal social benefits of most goods and services do not equal the marginal social costs. Part of the problem is that many of our actions have spillover effects on other people (these are known as 'externalities'), part is a lack of competition, part is a lack of knowledge by both producers and consumers, and part is the fact that markets may take a long time to adjust to any disequilibrium, given the often considerable short-run immobility of factors of production.

KEY IDEA 26

Markets generally fail to achieve social efficiency. There are various types of market failure. Market failures provide one of the major justifications for government intervention in the economy.

In this chapter we examine these various 'failings' of the free market and what the government can do to rectify the situation. We also examine why the government itself may fail to achieve social efficiency.

General equilibrium

Markets are in a constant state of flux. Demand changes as consumer tastes change and as income changes; supply changes as technology, the availability of natural resources and costs change. These changes in demand and supply cause markets to adjust to a new equilibrium.

At any one time, it is useful to look at the overall equilibrium towards which markets are heading: the **general equilibrium** of all markets.

The concept of general equilibrium is a threshold concept because it gives us an insight into how market forces apply to a whole economy, and not just to its individual parts. It is about seeing how the whole jigsaw fits together and how changes ripple throughout the economy.

KEY IDEA 27

General equilibrium. The situation where all individual markets in the economy are in equilibrium: in other words, where demand equals supply in all markets. If demand or supply changes in any market, there are likely to be ripple effects into other markets (i.e. for substitute or complementary goods, in both demand and supply), until a new general equilibrium is reached.

TC 13

We can then ask whether this general equilibrium is socially efficient, or whether a reallocation of resources in the economy would lead to greater social efficiency.

An economy where all markets are perfectly competitive and where there are no externalities *will* be socially efficient when there is a state of general equilibrium. Why will this be so? Take the case of goods markets. In any given goods markets, the consumer will achieve private efficiency where marginal utility (i.e. marginal benefit) equals price (see Section 4.1 on page 73); and the producer where marginal cost equals price (see Section 6.2, pages 124–5). Given that all producers and consumers face the same market price, in equilibrium marginal utility will equal marginal cost.

But will this be socially efficient? In the absence of externalities, benefits from consumption are confined to the consumers themselves. In other words, as members of society, their benefit is the whole social benefit. Thus $P = MU = MSB$. Likewise, the costs of production are confined to the producers: there are no costs imposed on *other* members of society. Thus $P = MC = MSC$.

To summarise:

$$MU = MSB = P = MC = MSC$$

and hence:

$$MSB = MSC$$

Thus equilibrium in any market under perfect competition with no externalities is socially efficient. When this applies to all markets, then general equilibrium is socially efficient.

Web Appendix 8.1 looks at social efficiency in more detail.

Definition

General equilibrium Where all the millions of markets throughout the economy are in a simultaneous state of equilibrium.

8.2 MARKET FAILURES: EXTERNALITIES AND PUBLIC GOODS

What will happen if certain markets are 'missing'?

Externalities

The analysis in Section 8.1 considers what happens under idealised conditions. However, the reality of the real world is that markets will fail to achieve social efficiency. One reason for this is that the actions of producers or consumers can affect people *other than themselves*. As we saw above, these effects on other people are known as **externalities**: they are the side-effects, or 'third-party' effects, of production or consumption.[1]

Externalities can be either desirable or undesirable. Whenever other people are affected beneficially, there are said to be **external benefits**. Whenever other people are affected adversely, there are said to be **external costs**.

KEY IDEA 28

Externalities are spillover costs or benefits. **Where these exist, even an otherwise perfect market will fail to achieve social efficiency.**

In the presence of externalities, the full benefit to society (the **social benefit**) from the consumption of any good is the private benefit enjoyed by consumers *plus* any externalities of consumption (positive or negative). Likewise, the full cost to society (the **social cost**) of the production of any good or service is the private cost faced by firms *plus* any externalities of production (positive or negative).

[1] Note that externalities are only those third-party effects *not* reflected in the market. So, for example, the effects of the actions of firms (e.g. price cuts) on their competitors would not class as externalities.

Definitions

Externalities Costs or benefits of production or consumption experienced by society but not by the producers or consumers themselves. Sometimes referred to as 'spillover' or 'third-party' costs or benefits.

External benefits Benefits from production (or consumption) experienced by people ***other*** than the producer (or consumer).

External costs Costs of production (or consumption) borne by people ***other*** than the producer (or consumer).

Social benefit Private benefit plus externalities in consumption.

Social cost Private cost plus externalities in production.

We now consider four different types of externality. In each case, we will assume that the market is in other respects perfect.

External costs of production (MSC > MPC) with no external costs/benefits of consumption (MSB = MPB)

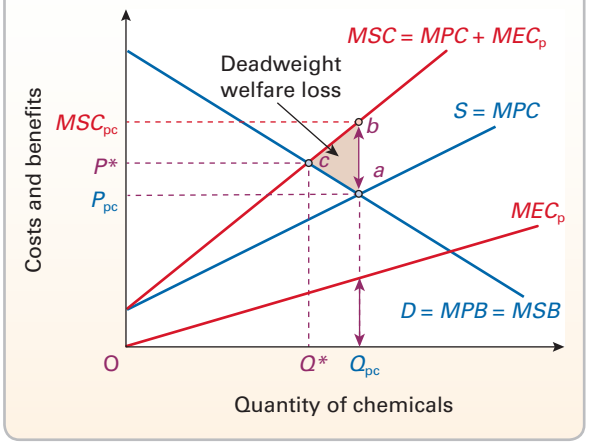
When a chemical firm dumps waste into a river or pollutes the air, the community bears costs additional to those borne by the firm. There are marginal external costs (MEC_p) of chemical production. This is illustrated in Figure 8.1. In this example we assume that they begin with the first unit of production and increase at a constant rate.

The marginal *social* costs (*MSC*) of chemical production will equal the marginal private costs (*MPC*) borne by the chemical firm plus the marginal external costs from production MEC_p This means that the *MSC* curve is above the *MPC* curve. The vertical distance between them is equal to the MEC_p. It is also assumed that there are no externalities in consumption, which means that the marginal social benefit (*MSB*) curve is the same as the marginal private benefit (*MPB*) curve.

Firms will maximise profit where marginal *private* cost (*MPC*) equals marginal revenue (*MR*). But under perfect competition, firms are price takers and thus can sell as much as they choose at the market price. Thus, for them, $P = AR = MR$ (see Section 6.2 on pages 124–5). This makes the *industry MPC* curve also the market supply curve, since at any price, firms will choose to supply the output where price equals marginal cost. Thus in Figure 8.1, $S = MPC$.

The market demand curve will be the sum of individuals' demand curves, which are equal to their

Figure 8.1 Negative externality in production



marginal utility curves (see Section 4.1 on pages 74–5). In the context of consumption, marginal utility is the same thing as marginal private benefit. As we are assuming there are no externalities on the consumption side, $D = MPB = MSB$.

Competitive market forces, with producers and consumers only responding to private costs and benefits, will result in a market equilibrium at point *a* in Figure 8.1: i.e. where demand equals supply. The market equilibrium price is P_{pc}, while the market equilibrium quantity is Q_{pc}.

At P_{pc}, *MPB* is equal to *MSB*. The market price reflects both the private and social benefits from the last unit consumed. However, the presence of external costs in production means that $MSC > MPC$.

The **socially optimal output** would be Q^*, where $P = MSB = MSC$. This is illustrated at point *c* and clearly shows how external costs of production in a perfectly competitive market result in overproduction i.e. $Q_{pc} > Q^*$. From society's point of view, too much waste is being dumped in rivers. Society therefore suffers a welfare loss.

We can show this welfare loss diagrammatically. If we compare the sum of the marginal social costs of units produced between Q_{pc} and Q^* with the marginal social benefits of these units we can see that this welfare loss is the area *abc*. This is the **deadweight welfare loss** of the overproduction. Therefore moving from Q_{pc} to Q^* would represent a *gain* in **social surplus** of the area *abc*.

One of the reasons why external costs cause problems in a free-market economy is because no one has legal ownership of the air or rivers and no one, therefore, can prevent or charge for their use as a dump for waste. Such a 'market' is *missing*. Control must, therefore, be left to the government, local authorities or regulators.

Other examples of external costs of production include extensive farming that destroys hedgerows and wildlife, and global warming caused by CO_2 emissions from power stations, industry and freight transport.

Definitions

Socially optimal output The output where $MSC = MSB$: the output where total social surplus is maximised.

Deadweight loss The loss in welfare (social surplus) arising from a socially inefficient allocation of resources.

Social surplus Total social benefits minus total social costs.

External benefits of production (MSC < MPC) with no external costs/benefits of consumption (MSB = MPB)

If companies in the forestry industry plant new woodlands, there is a benefit not only to the companies and their customers, but also to the world through a reduction of CO_2 in the atmosphere (forests are a carbon sink). In this case there are marginal external benefits (MEB_p) of production. These are shown in Figure 8.2. We assume that they begin with the first tree planted but that the marginal benefit declines with each additional tree. In other words, the MEB_p is a downward-sloping line.

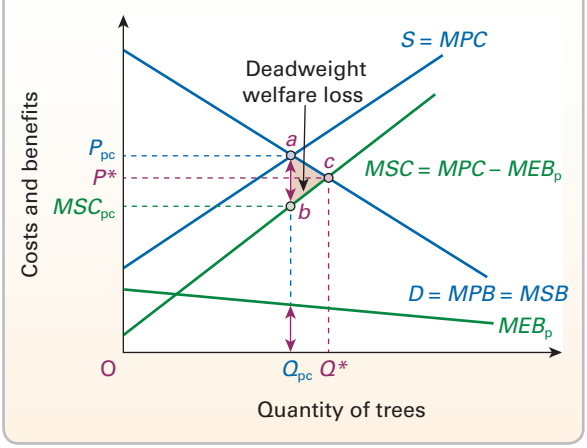
Pause for thought

Why are marginal external benefits typically likely to decline as output increases? Why in some cases might marginal external benefits be constant at all levels of output or even increase as more is produced?

Given these positive externalities, the marginal *social* cost (*MSC*) of providing timber is less than the marginal private cost: $MSC = MPC - MEB_p$. This means that the *MSC* curve is *below* the *MPC* curve. The vertical distance between the curves is equal to the MEB_p. Once again, it is assumed that there are no externalities in consumption so that $MSB = MPB$.

Competitive market forces will result in an equilibrium output of Q_{pc}, where market demand (=*MPB*) equals market supply (=*MPC*) (point *a*). The socially efficient level of output, however, is Q^*: i.e. where $MSB = MSC$ (point *c*). The external benefits of production thus result in a level of output *below* the socially efficient level. From society's point of view

Figure 8.2 Positive externality in production



not enough trees are being planted. The deadweight welfare loss caused by this underproduction is illustrated by the area *abc*. Output is not being produced between Q_{pc} and Q^* even though $MSB > MSC$.

Another example of external benefits in production is that of research and development. An interesting recent example has been the development of COVID-19 vaccines. If other firms have access to the results of the research, then clearly the benefits extend beyond the firm that finances it. Since the firm only receives the private benefits, it will conduct a less than optimal amount of research. In turn, this may reduce the pace of innovation and cause slower long-term economic growth.

External costs of consumption (MSB < MPB) with no external costs/benefits of production (MSC = MPC)

Drinking alcohol can sometimes lead to marginal external costs of consumption. For example, in 2019/20 there were 280 000 admissions to hospitals in England where the main reason was attributable to alcohol consumption. In 2020, approximately 6500 people were either killed or injured in an accident where at least one driver was over the drink-drive limit. These marginal external costs of consumption (MEC_c) result in the marginal social benefit of alcohol consumption being lower than the marginal private benefit: i.e. $MSB = MPB - MEC_c$

This is illustrated in Figure 8.3, where the *MSB* curve is below the *MPB* curve. In this example it is assumed that there are no externalities in production so that $MSC = MPC$.

Competitive market forces will result in an equilibrium output of Q_{pc} (point *a*) whereas the socially efficient level of output is Q^*: i.e. where $MSB = MSC$ (point c). The external costs of consumption result in level of output *above* the socially efficient level: i.e. $Q_{pc} > Q^*$. From society's point of view, too much alcohol is being produced and consumed. The deadweight welfare loss caused by this overconsumption is illustrated by the area *abc*.

Other possible examples of negative externalities of consumption include taking a journey by car, noisy radios in public places, the smoke from cigarettes and litter.

External benefits of consumption (MSB > MPB) with no external costs/benefits of production (MSC = MPC)

How do people travel to a city centre to go shopping on a Saturday? How do people travel to a football match? If they use the train, then other people benefit, as there is less congestion and exhaust fumes and fewer accidents on the roads. These marginal external benefits of consumption (MEB_c) result in the marginal social benefit of rail travel being *greater* than the marginal private benefit (i.e. $MSB = MPB + MEB_c$).

This is illustrated in Figure 8.4, where the *MSB* curve is above the *MPB* curve. The vertical distance between the curves is equal to the MEB_c. Once again it is assumed that there are no externalities in production so that $MSC = MPC$.

External benefits of consumption result in a level of output below the socially efficient level i.e. $Q_{pc} < Q^*$. From society's point of view not enough journeys are being made by train. The deadweight welfare caused by this underconsumption is illustrated by the area *abc*.

Other examples of external benefits of consumption include the beneficial effects for other people from someone having a vaccination, wearing a face mask, cycling, using a deodorant and planting flowers in their front garden.

To summarise: whenever there are external benefits, there will be too little produced or consumed.

TC 12 p 195

Figure 8.3 Negative externality in consumption

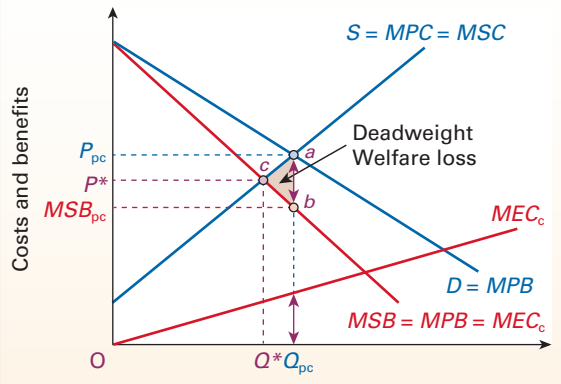
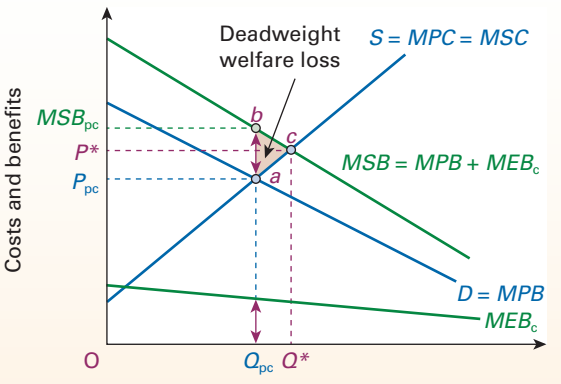


Figure 8.4 Positive externality in consumption



Whenever there are external costs, there will be too much produced or consumed. The market will not equate *MSB* and *MSC*.

The above arguments have been developed in the context of perfect competition, with prices determined by demand and supply. Externalities also occur in all other types of market.

Public goods

There is a category of goods where the positive externalities are so great that the free market, whether perfect or imperfect, may not produce at all. They are called **public goods**. Examples include lighthouses for private pleasure craft,[2] pavements, flood-control dams, public drainage, public services such as the police and even government itself.

Public goods have two important characteristics: *non-rivalry* and *non-excludability*.

- If I consume a bar of chocolate, it cannot then be consumed by someone else. If, however, I enjoy the benefits of street lighting, it does not prevent you or anyone else doing the same. There is thus what we call **non-rivalry** in the consumption of such goods. These goods tend to have large external benefits relative to private benefits. This makes them socially desirable, but privately unprofitable. No single individual would pay to have a pavement built along his or her street. The private benefit would be too small relative to the cost. And yet the social benefit to all the other people using the pavement may far outweigh the cost.

KI 28 p 196

KEY IDEA 29

The free-rider problem. **People are often unwilling to pay for things if they can make use of things other people have bought. This problem can lead to people not purchasing things that would be to the benefit of themselves and other members of society.**

- If I spend money erecting a flood-control dam to protect my house, my neighbours will also be protected by the dam. I cannot prevent them enjoying the benefits of my expenditure. This feature of **non-excludability** means that they would get the benefits free and would therefore have no incentive to pay themselves. This is known as the **free-rider problem**.

Pause for thought

Which of the following have the property of non-rivalry: (a) a can of drink; (b) public transport; (c) a commercial radio broadcast; (d) the sight of flowers in a public park?

When goods have these two features, the free market will simply not provide them. Thus these public goods can only be provided by the government or by the government subsidising private firms to do so. However, it is important to note that not all goods and services produced by the public sector come into the category of 'public goods and services': thus education and health are publicly provided, but they *can* be, and indeed are, privately provided as well.

Definitions

Public good A good or service that has the features of non-rivalry and non-excludability and as a result would not be provided by the free market.

Non-rivalry Where the consumption of a good or service by one person will not prevent others from enjoying it.

Non-excludability Where it is not possible to provide a good or service to one person without it thereby being available to others to enjoy.

Free-rider problem Where it is not possible to exclude other people from consuming a good that someone has bought.

[2] As far as commercial ships are concerned, it is practical to levy a charge. Merchant ships and larger commercial pleasure craft calling at UK and Irish ports are charged 'light dues', which are used to finance lighthouse services.

Recap

1. Social efficiency will be achieved where *MSC* = *MSB* for each good and service. In practice, however, markets fail to achieve social efficiency. One reason for this is the existence of externalities.
2. Externalities are spillover costs or benefits. Whenever there are external costs, the market will (other things being equal) lead to a level of production and consumption *above* the socially efficient level. Whenever there are external benefits, the market will (other things being equal) lead to a level of production and consumption *below* the socially efficient level.
3. Public goods will not be provided by a free market. The problem is that they have large external benefits relative to private benefits and without government intervention it would not be possible to prevent people having a 'free ride' and thereby escaping contributing to their cost of production.

BOX 8.1 **THE TRAGEDY OF THE COMMONS** EXPLORING ECONOMICS

The depletion of common resources

Common resources are not owned but are available free of charge to anyone. Examples include the air we breathe and the oceans for fishing. Like public goods, they are *non-excludable*. For example, in the absence of intervention, fishing boats can take as many fish as they are able to from the open seas. There is no 'owner' of the fish to stop them. As long as there are plentiful stocks of fish, there is no problem.

But as more people fish the seas, so fish stocks are likely to run down. This is where common resources differ from public goods. There *is* rivalry. One person's use of a common resource diminishes the amount available for others. This result is an overuse of common resources. This is why many fish stocks are severely depleted, why rainforests are disappearing (cut down for timber or firewood), why many roads are congested and why the atmosphere is so polluted (being used as a common 'dump' for emissions). In each case, a resource that is freely available is overused. This has become known as the **tragedy of the commons.**

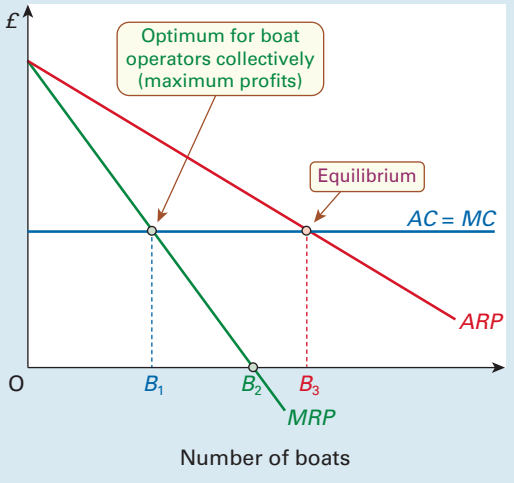
How can we analyse the overuse of common resources? The simplest way is in terms of externalities. When I use a common resource, I am reducing the amount available for others. I am imposing a cost on other people: an external cost. If I am motivated by self-interest, I will not take these external costs into account and so an overuse of the resource occurs.

KI 28 p 196

Another way of analysing it is to examine the effect of one person's use of a resource on other people's output externalities . Take the case of fishing grounds. In the diagram the horizontal axis measures the use of this common resource, say in terms of the number of fishing boats per day. The average cost of operating a boat (e.g. the wages of the crew and the fuel) is taken to be constant and is thus equal to the marginal cost. For the sake of simplicity, the price of fish is also assumed to be constant.

KI 2 p 7

Fishing in open-access fishing grounds



As the number of boats increases and fish stocks decline, so each extra boat entering will add less and less to the total catch. The revenue added by each extra boat is the marginal revenue product (*MRP*). In this case the additional collective revenue from each extra boat declines. Eventually, at point B_2, no more fish can be caught: $MRP = 0$. The catch is at the maximum. The average revenue product (*ARP*) is the revenue earned per boat: i.e. the total value of the catch divided by the number of boats.

The average and marginal revenue product curves have to be interpreted with care. Say one additional boat enters the fishing ground. The *MRP* curve shows the extra revenue accruing to the boat operators collectively. It does *not* show the revenue actually earned by the additional boat. The extra boat gets an average catch (which has been reduced somewhat because of the additional boat) and hence gains the average revenue product of all the boats.

What will be the equilibrium? Note first that the optimal number of boats for the boat operators collectively is B_1, where the marginal cost of an extra boat equals its marginal revenue product. In other words, this maximises the collective profit. At point B_1, however, there will be an incentive for extra boats to enter the fishery because the average revenue product (that is, the return that an additional boat gets) is greater than the cost of operating the boat.

More boats will enter as long as the value earned by each boat (*ARP*) is greater than the cost of operating it: as long as the *ARP* curve is above the $AC = MC$ line. Equilibrium is reached with B_3 boats: considerably above the collective profit-maximising number. Note also that the way the diagram is drawn, marginal revenue product is negative. The last boat has *decreased* the total value of the catch.

In many parts of the world, fish stocks have become so severely depleted that governments, individually or collectively, have had to act. Measures have included quotas on catches or the number of boats, minimum net mesh sizes (to allow young fish to escape), or banning fishing altogether in certain areas or for certain species.

1. To what extent can the following be regarded as common resources? (a) rainforests; (b) children's playgrounds in public parks; (c) silence in a library; (d) the Internet.

Nobel Prize winning economist Elinor Ostrom has discovered examples of real-world common resources that are consumed in a sustainable manner. Conduct a literature search and other desktop research and summarise Ostrom's work on common resources.

8.3 MARKET FAILURES: MONOPOLY POWER

What problems arise from big business?

KI 26 p 195

Whenever markets are imperfect, whether as pure monopoly or monopsony, or whether as some form of imperfect competition, the market will fail to equate *MSB* and *MSC*, even if there are no externalities.

Take the case of monopoly. A monopoly will produce less than the socially efficient output. This is illustrated in Figure 8.5. A monopoly faces a downward-sloping demand curve, and therefore marginal revenue is below average revenue (=P = MSB). Profits are maximised at an output of Q_1, where marginal revenue equals marginal cost (see Figure 6.5 on page 132). If there are no externalities, the socially efficient output will be at the higher level of Q_2, where $MSB = MSC$.

Deadweight loss under monopoly

Consumer and producer surplus

One way of analysing the welfare loss that occurs under monopoly is to use the concepts of *consumer* and *producer surplus*.

We first came across consumer surplus in Section 4.1 (see page 74). It is the excess of consumers' total benefit (or 'utility') from consuming a good over their total expenditure on it.

Producer surplus is similar to supernormal profit (see Section 5.4). The two concepts are illustrated in Figure 8.6. The diagram shows an industry that is initially under perfect competition and then becomes a monopoly (but faces the same revenue and cost curves).

Let us start by examining consumer and producer surplus under *perfect competition*.

Figure 8.5 The monopolist producing less than the social optimum

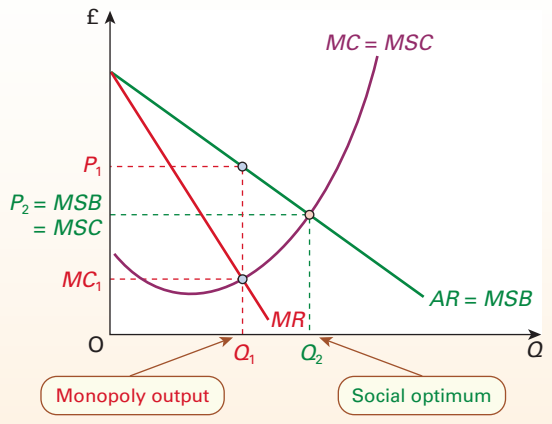
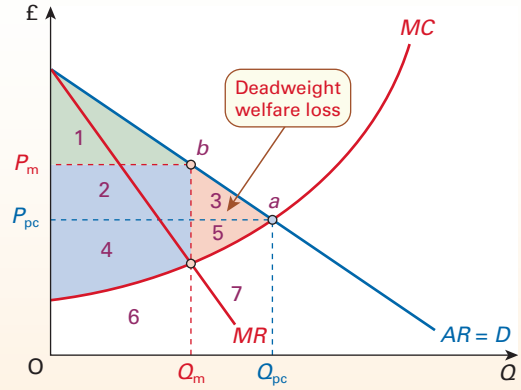


Figure 8.6 Deadweight loss from a monopoly



Consumer surplus. Under perfect competition the industry will produce an output of Q_{pc} at a price of P_{pc}, where $MC (=S) = P (=AR)$: i.e. at point *a* (see pages 132–3).

Consumers' total benefit is given by the area under the demand curve (the sum of all the areas 1–7). The reason for this is that each point on the demand curve shows how much the last consumer is prepared to pay (i.e. the benefit to the marginal consumer). The area under the demand curve thus shows the total of all these marginal benefits from zero consumption to the current level: i.e. it gives total benefit.

Consumers' total expenditure is $P_{pc} \times Q_{pc}$ (areas 4 + 5 + 6 + 7).

Consumer surplus is the difference between total benefit and total expenditure: in other words, the area between the price and the demand curve (areas 1 + 2 + 3).

Producer surplus. **Producer surplus** is the difference between total revenue and total variable cost.

Total variable cost is the area under the *MC* curve (areas 6 + 7). The reason for this is that each point on the marginal cost curve shows what the last unit costs

Definitions

Consumer surplus The excess of what a person would have been prepared to pay for a good (i.e. the utility) over what the person actually pays (see also definitions on page 74)

Producer surplus The excess of total revenue over total variable cost.

to produce. The area under the *MC* curve thus gives all the marginal costs starting from an output of zero to the current output: i.e. it gives total variable costs.

Total revenue is $P_{pc} \times Q_{pc}$ (areas 4 + 5 + 6 + 7).

Producer surplus is thus the area between the price and the *MC* curve (areas 4 + 5).[3]

Total (private) surplus. Total consumer plus producer surplus is therefore the area between the demand and *MC* curves. This is shown by the total shaded area (areas 1 + 2 + 3 + 4 + 5).

The effect of monopoly on total surplus

What happens when the industry is under *monopoly*? The firm will produce where $MC = MR$, at an output of Q_m and a price of P_m (at point *b* on the demand curve). Total revenue is $P_m \times Q_m$ (areas 2 + 4 + 6). Total cost is the area under the *MC* curve (area 6). Thus producer surplus is areas 2 + 4. This is clearly a *larger* surplus than under perfect competition (since area 2 is larger than area 5): monopoly profits are larger than profits under perfect competition.

KI 19 p 122

Consumer surplus, however, will be much smaller. With consumption at Q_m, total benefit to consumers is given by areas 1 + 2 + 4 + 6, whereas consumer expenditure is given by areas 2 + 4 + 6. Consumer surplus, then, is simply area 1. (Note that area 2 has been transformed from consumer surplus to producer surplus.)

Total surplus under monopoly is therefore areas 1 + 2 + 4: a smaller surplus than under perfect competition. 'Monopolisation' of the industry has resulted in a loss of total surplus of areas 3 + 5. The producer's gain is less than consumers' loss. This net loss of total surplus is the deadweight welfare loss of monopoly.

KI 26 p 195

[3] As the sum of marginal costs gives total variable costs, it does not include *fixed* costs. Total profit, therefore, is less than total producer surplus. Total profit is total producers' surplus minus total fixed costs.

Conclusions

As was shown in Section 6.3, there are possible social *advantages* from powerful firms: advantages such as economies of scale and more research and development. These advantages may outweigh deadweight loss from monopoly power. It can be argued that an ideal situation would be where firms are large enough to gain economies of scale and yet were somehow persuaded or compelled to produce where $P = MC$ (assuming no externalities).

Pause for thought

Assume that a monopoly existed in an industry where there were negative externalities. Could the socially efficient output be Q_m in Figure 8.6? If so, would this make monopoly socially efficient?

Recap

1. Monopoly power will (other things being equal) lead to a level of output below the socially efficient level.
2. This will result in deadweight welfare loss, which is the loss in total producer and consumer surplus.
3. Consumer surplus is the excess of what consumers are prepared to pay (which is how we measure the benefit to consumers) over what they actually pay. Producer surplus is the excess of total revenue over total cost (i.e. total profit).
4. The effect of monopoly will be to give a higher producer surplus than under perfect competition, but a much lower consumer surplus. Thus total surplus is lower.
5. There are potential gains from monopoly, such as economies of scale and higher investment. Such gains have to be offset against the deadweight loss.

8.4 OTHER MARKET FAILURES

In what other ways may a market fail to make the best use of scarce resources?

Imperfect information

Ignorance and uncertainty

Perfect competition assumes that consumers, firms and factor suppliers have perfect knowledge of costs and benefits. In the real world there is often a great deal of ignorance and uncertainty. Thus people are unable to equate marginal benefit with marginal cost.

Consumers purchase many goods infrequently. Examples include houses, cars, washing machines, televisions and other consumer durables. Consumers may not be aware of the quality of such goods until

they have purchased them, by which time it is too late. Advertising may contribute to people's ignorance by misleading them as to the benefits of a good.

Firms are often ignorant of market opportunities, prices, costs, the productivity of labour (especially white-collar workers), the activity of rivals, etc.

Many economic decisions are based on expected future conditions. Since the future can never be known for certain, many decisions may turn out to be wrong.

TC 13 p 195

In some cases it may be possible to obtain the information through the market. There may be an agency that will sell you the information, or a newspaper, magazine or website that contains the information. In such cases you will have to decide whether the cost to you of obtaining the information is worth the benefit it will provide you. A problem here is that you may not have sufficient information to judge how reliable the information is that you are obtaining!

Pause for thought

1. *Make a list of pieces of information that a firm might want to know, and consider whether it could buy the information and how reliable that information might be.*
2. *What has been the impact of the Internet on the provision of information?*

Asymmetric information

One form of imperfect information occurs when different sides in an economic relationship have different amounts of information. As we saw in Section 4.3 (page 83) this is known as **asymmetric information** and is at the heart of the *principal–agent problem* (see Section 5.5 pages 116–17).

KI 18 p 117

The danger is that the agent may well not act in the principal's best interests and may be able to get away with it because of the principal's imperfect knowledge. The estate agent trying to sell you a house may not tell you about the noisy neighbours or that the vendor is prepared to accept a much lower price. A second-hand car dealer may 'neglect' to tell you about the rust on the underside of the car or that it has a history of unreliability. A financial institution might recommend loans to a borrower at fixed rates of interest, knowing that rates were likely to fall resulting in the borrower paying a higher rate of interest than if they had taken out a variable rate loan.

Definition

Asymmetric information Where one party in an economic relationship (e.g. an agent) has more information than another (e.g. the principal).

Immobility of factors and time lags in response

Even under conditions of perfect competition, factors may be very slow to respond to changes in demand or supply. Labour, for example, may be highly immobile both occupationally and geographically. This can lead to large price changes and hence to large supernormal profits and high wages for those in the sectors of rising demand or falling costs. The long run may be a very long time coming!

In the meantime, there will be further changes in the conditions of demand and supply. Thus the economy is in a constant state of disequilibrium and the long run never comes. As firms and consumers respond to market signals and move towards equilibrium, so the equilibrium position moves, and the social optimum is never achieved.

KEY IDEA 30

The problem of time lags. Many economic actions can take a long time to take effect. This can cause problems of instability and an inability of the economy to achieve social efficiency.

Whenever monopoly/monopsony power exists, the problem is made worse as firms or unions put up barriers to the entry of new firms or factors of production.

Protecting people's interests

Dependants

People do not always make their own economic decisions. They are often dependent on decisions made by others. Parents make decisions on behalf of their children; partners on each other's behalf; younger adults on behalf of old people; managers on behalf of shareholders, etc. This, again, is an example of the principal–agent problem.

KI 18 p 117

A free market will respond to these decisions, however good or bad they may be, and whether or not they are in the interests of the dependants. Thus the government may feel it necessary to protect those whose welfare depends on choices made by others.

Poor economic decision-making by individuals on their own behalf

The government may feel that people need protecting from poor economic decisions that they make on their *own* behalf. As we discussed in Box 4.3 (see page 79) this may be a particular problem when the benefits from consuming a good are immediate while the cost happens at some point in the future. People may place too much weight on the immediate benefits and too little weight on the long-run costs of their decisions. Examples might include the consumption of alcohol, tobacco and fast/unhealthy food.

BOX 8.2 SHOULD HEALTH-CARE PROVISION BE LEFT TO THE MARKET?

A case of multiple market failures

In the UK, the National Health Service provides free hospital treatment, a free general practitioner service and free prescriptions for certain categories of people. Their marginal cost to the patient is thus zero. Of course, these services use resources and they thus have to be paid for out of taxes.

But why are these services not sold directly to the patient, thereby saving the taxpayer money? There are, in fact, a number of reasons why the market would fail to provide the optimum amount of health care.

The issue of equity

KI 24 p 185

This is a problem connected with the distribution of income. Because income is unequally distributed, some people will be able to afford better treatment than others, and the poorest people may not be able to afford treatment at all. On grounds of equity, therefore, it is argued that health care should be provided free – at least for those on low incomes.

The concept of equity that is usually applied to health care is that individuals should be able to access treatment according to their medical need rather than according to their ability to pay. This was brought into sharp relief with the COVID-19 pandemic. It was generally believed that access to NHS resources should depend solely on clinical need.

Difficulty for people in predicting their future medical needs

KI 15 p 80

If you were suddenly taken ill and required a major operation, it could be very expensive indeed for you if you had to pay. On the other hand, you may go through life requiring very little if any medical treatment. In other words, there is great uncertainty about your future medical needs. As a result, it would be very difficult to plan your finances and budget for possible future medical expenses if you had to pay for treatment.

Medical insurance could provide a solution to this problem. However, as discussed in Box 4.5 this market can be affected by problems of adverse selection (people who take out insurance are likely to be those who have the highest risk) and moral hazard (having insurance makes you less careful and thus increases your risk to the company).

Even if an effective market for medical insurance could be established there remains a problem of equity. Would the chronically sick or very old be able to obtain cover and, if so, would they be able to afford the premiums? This issue of 'gaps' in an insurance-based system, means that some form of intervention on grounds of equity may be needed, even if most provision is private.

Externalities

KI 28 p 196

Health care generates a number of benefits *external* to the patient. If you are cured of an infectious disease, for example, it is not just you who benefits but also others, since you will not infect them. In addition, if you have a job you will be able to get back to work, thus reducing the disruption there. These external benefits of health care could be quite large.

If the sick had to pay the cost of their treatment, they may decide not to be treated – especially if they are poor. They will consider the costs and benefits they might experience, but may not take into account the effect that their illness has on other people. The market, by equating *private* benefits and costs, would produce too little health care.

Information problems and patient ignorance

Markets only function well to allocate resources efficiently if the consumer has the information to make informed decisions. For many products we purchase, there is reasonably good information and so we can judge which products/services we will like the most.

In the case of health care, the situation is different. Much of the information is complex and a patient might require specialist knowledge to understand it. The one-off nature of

Therefore, if the government wanted to discourage smoking and drinking, for example, it could put taxes on tobacco and alcohol to try to affect the choices that people make. In more extreme cases, it could make various activities illegal: activities such as driving over the limit, certain types of gambling, the sale and consumption of drugs and attending parties during COVID lockdowns.

On the other hand, the government may feel that people consume too little of things that are good for them: things such as education, health care and sports facilities. Such goods are known as **merit goods**. The government could provide them free, subsidise their production or make their consumption compulsory (e.g. as with fluoride added to water in many areas to protect people's teeth).

Pause for thought

How do merit goods differ from public goods?

Other objectives

As we saw in Chapter 7, one of the major criticisms of the free market is the problem of *inequality*. An efficient allocation of resources can in theory be achieved for any *given* distribution of income. Hence, both an equal and a highly unequal distribution of income could be consistent with social efficiency and

Definition

Merit goods Goods which the government feels that people should consume but tend to underconsume and which therefore ought to be subsidised or provided free.

CASE STUDIES & APPLICATIONS

many treatments also means that patients are unable to learn from repeat purchases.

In some circumstances the demand for health care might be urgent. In these cases people do not have the time to shop around and may be unable to make rational decisions because of heightened emotional distress. For these reasons patients have to rely on the professional advice of others. They have to enter an agency relationship with a doctor or health-care provider – patients are the principals; heathcare professionals are the agents.

For example, if you have a pain in your chest, it may be simple muscular strain, or it may be a symptom of heart disease. You rely on the doctor (the *supplier* of the treatment) to give you the information: to diagnose your condition. The key issue is whether the incentives of the doctor and the patient are aligned.

KI 18 p 117

If health care was provided through the market, unscrupulous doctors might advise more expensive treatment than is necessary; they might even have an agreement with certain drugs companies that they will try to persuade you to buy an expensive branded product rather than an identical cheaper version. The problem will also exist in an insurance-based system, where the doctor may be even more inclined to over-supply if the patient has sufficient cover.

KI 18 p 117

If people had to pay, those patients suffering from the early stages of a serious disease might not consult their doctor until the symptoms become acute, by which time it might be too late to treat the disease, or very expensive to do so. With a health service that is free at the point of use, however, a person is likely to receive an earlier diagnosis of serious conditions.

Oligopoly

If doctors and hospitals operated in the free market as profit maximisers, it is unlikely that competition would drive down their prices. Instead they might collude to fix standard prices for treatment, so as to protect their incomes. Even if doctors did compete openly, it is unlikely that consumers would have enough information to enable them to 'shop around' for the best value.

KI 19 p 122

We have to be careful though: to argue that the market system will fail to provide an optimal allocation of health-care resources does not in itself prove that *free provision* will result in optimal provision. For example, with no charge for GP appointments it is likely that some patients will consult their doctors over trivial complaints. The result will be excessive consumption, with consumption beyond the socially efficient point.

In the USA there is much more reliance on *private medical insurance*. Alternatively, the government may simply *subsidise* health care, so as to make it cheaper rather than free. This is the case with prescriptions and dental treatment in the UK, where many people have to pay part of the cost of treatment. Also, the government can *regulate* the behaviour of the providers of health care, to prevent exploitation of the patient. Thus only people with certain qualifications are allowed to operate as doctors, nurses, pharmacists, etc.

1. *Does the equity argument that individuals should be able to access medical treatment according to their medical need rather than according to their ability to pay also apply to food and other basic goods?*
2. *If health care is provided free, the demand is likely to be high. How is high demand likely to be dealt with? Is this a good way of dealing with the issue?*

Go through each of the market failings identified in this box. In each case, write a short summary of the alternative policies open to a government to tackle them and discuss their advantages and disadvantages.

KI 24 p 185

yet it could be argued that a more equal distribution is socially more desirable.

In addition to social efficiency and greater equality, we can identify other social goals: goals such as moral behaviour (however defined), enlightenment, social consciousness, co-operation, the development of culture, fulfilment, freedom from exploitation, and freedom to own, purchase and inherit property. The unfettered free market may not be very successful in achieving social efficiency. It may be even less successful in achieving many other social goals.

Finally, there are *macroeconomic objectives* which the free market is unlikely to achieve simultaneously. These relate in one way or another to the total level of spending or the total level of output in an economy. These objectives include strong and stable rates of economic growth, full employment, stable prices, a balance of international payments and a stable financial system.

Parts C and D of this text examine the various macroeconomic problems and the methods of government intervention in more detail.

How far can economists go in advising governments?

Certain goals, such as growth in national income, changes in the distribution of income and greater efficiency, are relatively easy to quantify. Others, such as enlightenment or the sense of community well-being, are virtually impossible to quantify. For this reason, economics tends to concentrate on the means of achieving a relatively narrow range of goals. The danger is that by economists concentrating on a limited number of goals, they may well influence the policy makers – the government, local authorities, various pressure groups, etc. – into doing the same, and thus into neglecting other perhaps important social goals.

Different objectives are likely to conflict. For example, economic growth may conflict with greater equality. In the case of such 'trade-offs', all the economist can do is to demonstrate the effects of a given policy, and leave the policy makers to decide whether the benefits in terms of one goal outweigh the costs in terms of another goal.

TC 1 p 7

KEY IDEA 31

Societies face trade-offs between economic objectives. **For example, the goal of faster growth may conflict with that of greater equality; the goal of lower unemployment may conflict with that of lower inflation (at least in the short run). This is an example of opportunity cost: the cost of achieving more of one objective may be achieving less of another. The existence of trade-offs means that policy makers must make choices.**

Recap

1. Ignorance and uncertainty may prevent people from consuming or producing the levels they would otherwise choose. Information may sometimes be provided (at a price) by the market, but it may be imperfect and, in some cases, not available at all.
2. Asymmetric information may result in the principal–agent problem, where lack of information by principals may result in the agents not operating in the principals' best interests.
3. Markets may respond sluggishly to changes in demand and supply. The time lags in adjustment can lead to a permanent state of disequilibrium and to problems of instability.
4. In a free market there may be inadequate provision for dependants and an inadequate output of merit goods; there are likely to be macroeconomic problems and problems of inequality and poverty; finally, there may be a whole series of social, moral, attitudinal and aesthetic problems arising from a market system.
5. Although economists cannot make ultimate pronouncements on the rights and wrongs of the market – that involves making moral judgements (and economists here are no different from any other person) – they can point out the consequences of the market and of various government policies, and also the trade-offs that exist between different objectives.

8.5 GOVERNMENT INTERVENTION: TAXES AND SUBSIDIES

Will taxing the bad and subsidising the good solve the problem of externalities?

Faced with all the problems of the free market, what is a government to do?

There are several policy instruments that the government can use. At one extreme it can totally replace the market by providing goods and services itself. At the other extreme it can merely seek to persuade producers, consumers or workers to act differently. Between the two extremes the government has a number of instruments it can use to change the way markets operate. These include taxes, subsidies, laws and regulatory bodies. In this and the next two sections we examine these different forms of government intervention.

The use of taxes and subsidies

A policy instrument particularly favoured by many economists is that of taxes and subsidies. They can be used for two main purposes: (a) to promote greater social efficiency by altering the composition of production and consumption, and (b) to redistribute incomes. We examined their use for the second purpose in the last chapter. Here we examine their use to achieve greater social efficiency.

When there are imperfections in the market, social efficiency will not be achieved. Marginal social benefit (*MSB*) will not equal marginal social cost (*MSC*). A different level of output would be more desirable. Taxes and subsidies can be used to correct these imperfections. Essentially the approach is to tax those goods or activities where the market produces too much, and subsidise those where the market produces too little.

KI 25 p 195

Taxes and subsidies to correct externalities

The rule here is simple: the government should impose a tax equal to the marginal external cost (or grant a subsidy equal to the marginal external benefit). This is known as a **Pigouvian tax** (or **Pigouvian subsidy**) named after the economist Arthur Pigou.

KI 25 p 195

Definition

Pigouvian tax (or subsidy) A tax (or subsidy) designed to 'internalise' an externality. The marginal rate of a Pigouvian tax (or subsidy) should be equal to the marginal external cost (or benefit).

Previously we examined the impact of external costs of pollution created by the chemical *industry* as a whole. We will now focus on one *firm* within that industry, which otherwise is perfectly competitive. Our firm is thus a price taker. Assume that this particular chemical company emits smoke from a chimney and thus pollutes the atmosphere. This creates external costs for the people who breathe in the smoke. The marginal social cost of producing the chemicals thus exceeds the marginal private cost to the firm: $MSC > MC$.

KI 28 p 196

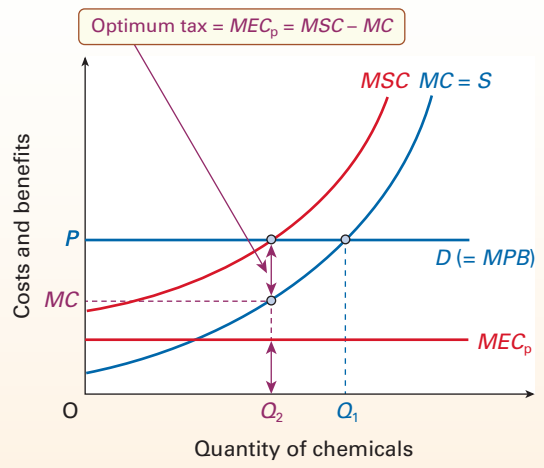
This is illustrated in Figure 8.7. In this example it is assumed the marginal external pollution cost begins with the first unit of production but remains constant. Hence the MEC_p is drawn as a horizontal line. The vertical distance between the *MC* and *MSC* curves is equal to the MEC_p. The marginal pollution cost (the externality) is shown by the vertical distance between the *MC* and *MSC* curves. The firm produces Q_1 where $P = MC$ (its profit-maximising output), but in doing so takes no account of the external pollution costs it imposes on society.

If the government now imposes a tax on production equal to the marginal pollution cost, it will effectively *internalise* the externality. The firm will have to pay an amount in tax equal to the external cost it creates. It will therefore now maximise profits at Q_2, where $P = MC$ + tax. But this is the socially optimum output where $MSB = MSC$.

TC 12 p 195

In the case of a firm producing an external *benefit*, it should be given a subsidy equal to that marginal external benefit. As is the case with a Pigouvian tax, this Pigouvian subsidy ought to be directed as closely as possible to the *source* of the externality. Take the case of a positive external benefit from a firm training its workers. Here the firm ought to be given a subsidy for each person trained rather than a general output subsidy. After all, an output subsidy would not only encourage the firm to train more people (the desired effect), but also encourage it to use more capital and raw materials (an undesired side-effect).

Figure 8.7 Using taxes to correct a distortion: an individual firm



Taxes and subsidies to correct for monopoly

If the problem of monopoly that the government wishes to tackle is that of *excessive profits*, it can impose a **lump-sum tax** on the monopolist: that is, a tax of a fixed absolute amount irrespective of how much the monopolist produces, or the price it charges. The use of a lump-sum tax is illustrated in Figure 8.8.

Since a lump-sum tax is an additional *fixed* cost to the firm, it will not affect the firm's marginal cost. Given the profit-maximising rule $MR = MC$, a lump-sum tax does not reduce the amount that the monopolist produces (which *would* be the case with a per-unit tax). Profits continue to be maximised at an output of Q_1 and a price of P_1. However, because the *AC* curve has shifted upwards, its profits are reduced from areas 1 + 2 to just area 1. Area 2 now represents the amount of tax paid to the government.

Pause for thought

How large would the lump-sum tax need to be for all supernormal profits to be eliminated?

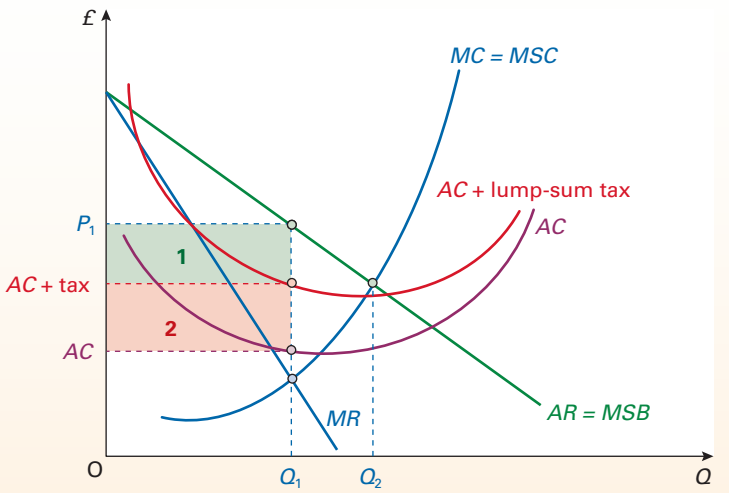
Two examples of lump-sum taxes are the 'windfall tax' imposed in 1997 by the UK Labour government on the profits of various privatised utilities, and the calls in 2022 for a windfall tax on oil and gas companies in the North Sea. This was in response to the rapid increase in their profits caused by the dramatic rise in global energy prices.

If the government is concerned that the monopolist produces less than the socially efficient output (Q_2 in Figure 8.8), it could give the monopolist a *per-unit subsidy* (which would encourage the monopolist to produce more). On its own, this would move both the *MC* and *AC* curves downwards. Since this would *increase* the monopolist's profit the government may wish to impose a harsh lump-sum tax in addition to the subsidy – which works to move the *AC* curve upwards. The lump-sum tax would not undo the subsidy's benefit of encouraging the monopolist to

Definition

Lump-sum tax A tax of a fixed amount independent of the circumstances of the entity being taxed (e.g. independent of the output of a firm or the income of an individual).

Figure 8.8 Using a lump-sum tax to reduce monopoly profits



produce more (it would not affect the *MC* curve), but it could be used to reduce the monopolist's profits below the pre-subsidy level.

Assessing the use of taxes and subsidies

Advantages of taxes and subsidies

Many economists favour the tax/subsidy solution to market imperfections (especially the problem of externalities) because it still allows the market to operate. It forces firms to take on board the full social costs and benefits of their actions. It is also adjustable according to the magnitude of the problem.

Moreover, if firms are taxed for polluting, they are encouraged to find cleaner ways of producing. The tax acts as an incentive over the longer run to reduce pollution. Likewise, by subsidising *good* practices, firms are given the incentive to adopt more good practices.

TC 6 p 24

Disadvantages of taxes and subsidies

Unfeasible to use different tax and subsidy rates. Each firm produces different levels and types of externalities and operates under different degrees of imperfect competition. It would be administratively very difficult and expensive, if not impossible, to charge every offending firm its own particular tax rate (or grant every relevant firm its own particular rate of subsidy). Even in the case of pollution where it is possible to measure a firm's emissions, there would still have to be a different tax rate for each pollutant and even for each environment, depending on its ability to absorb the pollutant and the number of people affected.

Using combinations of lump-sum taxes and per-unit subsidies to correct monopoly distortions to price, output and profit would also probably be impractical. Given that cost and revenue curves differ substantially from one firm to another, separate tax and subsidy rates would be needed for each firm. An army of tax inspectors would be necessary to administer the system!

Lack of knowledge. Even if a government did decide to charge a tax equal to each offending firm's marginal external costs, it would still have the problem of measuring those costs and apportioning blame. The damage to lakes and forests from acid rain has been a major concern since the beginning of the 1980s. But just how serious is that damage? What is its current monetary cost? How long-lasting is the damage?

Just what and who are to blame? These are questions that cannot be answered precisely. It is thus impossible to fix the 'correct' pollution tax on, say, a particular coal-fired power station.

Pause for thought

Why is it easier to use taxes and subsidies to tackle the problem of car exhaust pollution than to tackle the problem of peak-time traffic congestion in cities?

Despite these problems, it is nevertheless possible to charge firms by the amount of a particular emission. For example, firms could be charged for chimney smoke by so many parts per million of a given pollutant. Although it is difficult to 'fine-tune' such a system so that the charge reflects the precise number of people affected by the pollutant and by how much, it does go some way to internalising the externality. As Box 8.3 on page 220 shows, many countries in recent years have introduced 'green' taxes, seeing them as an effective means of protecting the environment.

Recap

1. Taxes and subsidies can be used to correct market distortions.
2. Externalities can be corrected by imposing tax rates equal to the size of marginal external costs, and granting rates of subsidy equal to marginal external benefits.
3. Taxes and subsidies can also be used to affect monopoly price, output and profit. Subsidies can be used to persuade a monopolist to increase output (and reduce price) to the competitive level. Lump-sum taxes can then be used to reduce monopoly profits without affecting the new price or output.
4. Taxes and subsidies have the advantages of 'internalising' externalities and of providing incentives to reduce external costs. On the other hand, they may be impractical to use when different rates are required for each case, or when it is impossible to know the full effects of the activities that the taxes or subsidies are being used to correct.

8.6 GOVERNMENT INTERVENTION: LAWS AND REGULATION

Should the government try to stop 'bad behaviour' by big business?

Laws prohibiting or regulating undesirable structures or behaviour

Laws are frequently used to correct market imperfections. Laws can be of three main types: those that prohibit or regulate behaviour that imposes external costs, those that prevent firms providing false or misleading information, and those that prevent or regulate monopolies and oligopolies.

Advantages of legal restrictions

- They are usually simple and clear to understand and are often relatively easy to administer. For example, various polluting activities could be banned or restricted.
- When the danger is very great, or when the extent of the danger is not as yet known, it might be much safer to ban various practices altogether (e.g. the use of various toxic chemicals) rather than to rely on taxes.
- When a decision needs to be taken quickly, it might be possible to invoke emergency action. For example, in a city it would be simpler to ban or restrict the use of private cars during a chemical smog emergency than to tax their use.
- Because consumers suffer from imperfect information, consumer protection laws can make it illegal for firms to sell shoddy or unsafe goods, or to make false or misleading claims about their products.

KI 18 p 117

Disadvantages of legal restrictions

The main problem is that legal restrictions tend to be a rather blunt weapon. If, for example, a firm were required to reduce the effluent of a toxic chemical to 20 tonnes per week, there would be no incentive for the firm to reduce it further. With a tax on the effluent, however, the more the firm reduced the effluent, the less tax it would pay. Thus with a system of taxes there is a *continuing* incentive to cut pollution, to improve safety, or whatever.

Regulatory bodies

Rather than using the blunt weapon of general legislation to ban or restrict various activities, a more 'subtle' approach can be adopted. This involves the use of various regulatory bodies. Having identified possible cases where action might be required (e.g. potential cases of pollution or the abuse of monopoly power), the regulatory body would probably conduct an investigation and then prepare a report containing its findings and recommendations. It might also have the power to enforce its decisions or this might be up to some higher authority.

EU competition policy

In most industrial countries, governments have in place a 'competition policy' which attempts to control the worst abuses of market power. Relevant EU legislation is contained in Articles 101 and 102 of the 2009 Treaty of the Functioning of the European Union. Additional regulations covering mergers came into force in 1990 and were amended in 2004. Further minor amendments have been put in place since then, which have focused on specific market regulation.

Restrictive practices. Article 101 is concerned with oligopolistic collusion. It is designed to prevent collusive *behaviour* rather than oligopolistic *structures* (i.e. the simple existence of co-operation between firms).

Practices considered anti-competitive include firms colluding to do any of the following:

- Fix prices (i.e. above competitive levels),
- Limit production, markets, technical development or investment,
- Share out markets or sources of supply,
- Charge discriminatory prices or operate discriminatory trading conditions, such as to benefit the colluding parties and disadvantage others,
- Make other firms who sign contracts with any of the colluding firms accept unfavourable obligations which, by their nature, have no connection with the subject of such contracts.

If companies are found guilty of undertaking any of these anti-competitive practices that are in contravention of Article 101, they are subject to financial penalties.

Abuse of market power. Article 102 relates to the abuse of market power and has also been extended to cover mergers. As with Article 101, it is the *behaviour* of firms that is the target of the legislation. The following are cited as examples of abuses of market power. As you can see, they are very similar to those in Article 101.

- Charging unfairly high prices to consumers, or paying unfairly low prices to suppliers.
- Limiting production, markets or technical developments to the detriment of consumers.
- Using price discrimination or other discriminatory practices to the detriment of certain parties.
- Making other firms that sign contracts with it accept unfavourable obligations which, by their nature, have no connection with the subject of such contracts.

Under Article 102, such practices can be banned and firms can be fined where they are found to have abused a dominant position.

The two Articles predominately apply to firms trading *between* EU members and so do not cover monopolies or oligopolies operating solely within a member country.

Merger policy. Under current regulations (2004), mergers that would significantly reduce competition in the EU are prohibited. This normally, therefore, applies to large mergers. For example, a merger could be blocked if there were concerns that the new firm would have significant market power that might lead to higher prices for consumers.

The merger must also have an 'EU dimension'. This is defined as where no more than two-thirds of each firm's EU-wide business is conducted in a single member state. If a firm does conduct more than two-thirds of its business in one country, then investigation of the merger would be the responsibility of that member state's competition authority.

UK competition policy

There have been substantial changes to UK competition policy since the first legislation was introduced in 1948. The current approach is based on the 1998 Competition Act and the 2002 Enterprise Act, together with Part 3 of the 2013 Enterprise and Regulatory Reform Act.

The Competition Act brought UK policy in line with EU policy, detailed above. The Act has two key sets (or 'chapters') of prohibitions. Chapter I prohibits various restrictive practices, and mirrors EU Article 101. Chapter II prohibits various abuses of monopoly power, and mirrors EU Article 102. The Enterprise Act strengthened the Competition Act and introduced new measures for the control of mergers.

The 2013 Act resulted in the setting up of a new regulatory body, the Competition and Markets Authority (CMA), to carry out investigations into particular firms or markets suspected of not working in the best interests of consumers and being in breach of one or more of the Acts. The CMA carries out much of the work previously undertaken by its predecessors, the Office for Fair Trading and the Competition Commission. Firms affected by a CMA ruling have the right of appeal to the independent Competition Appeal Tribunal (CAT), which can uphold or overturn the ruling.

Restrictive practices. The 1998 Competition Act brought UK restrictive practices policy more into line with EU policy. For example, the way fines were calculated and implemented for anti-competitive behaviour was changed so that it was more comparable to the method used by the European Commission: i.e. the penalties imposed could be up to 10 per cent of the firm's annual turnover.

The types of practices that constitute 'cartel agreements' were also made more consistent with EU policy: e.g. price fixing, limiting supply, sharing out markets, limiting supply or bid-rigging.

It is a *criminal* offence to engage in cartel agreements (i.e. horizontal, rather than vertical, collusive agreements between firms), irrespective of whether there are appreciable effects on competition. Convicted offenders can receive a prison sentence of up to five years and/or an unlimited fine. Prosecutions can be brought by the Serious Fraud Office or the CMA. The CMA also has substantial powers to enter premises, seize documents and require people to answer questions or provide information.

Abuse of market power. Under the Chapter II prohibition of the 1998 Competition Act, it is illegal for a dominant firm to exercise its market power in such a way as to reduce competition. Any suspected case is investigated by the CMA, which uses a two-stage process in deciding whether an abuse has taken place.

The first stage is to establish whether a firm has a position of dominance, where 'dominance' normally involves the firm having at least a 40 per cent share of the market. If the firm *is* deemed to be dominant, the second stage involves the CMA having to decide whether the firm's practices constitute an abuse of its position. It specifies the same four types of market abuse as does the EU's Article 102 (see above).

The simple *existence* of any of these practices may not constitute an abuse. The CMA has to decide whether their *effect* is to restrict competition. This may require a detailed investigation to establish whether competition is restricted or distorted. If this is found to be the case, the CMA decides what actions must be taken to remedy the situation.

Merger policy. A merger or takeover is investigated by the CMA if the target company has a UK turnover that exceeds £70 million, or if the merger results in the new company having a market share of 25 per cent or more.

UK policy is similar to EU policy. The CMA conducts a preliminary or Phase 1 investigation to see whether competition is likely to be threatened. There is a statutory deadline of 40 working days to complete Phase 1, by which point the CMA has to decide whether there is a significant chance that the merger would result in a substantial lessening of competition (SLC). If the CMA concludes that this might be the case, it begins Phase 2 of the process – a much more in-depth assessment. If no SLC issues are raised, the merger is allowed to go ahead.

There is a 24-week statutory time limit for Phase 2 decisions to be made, by which point the CMA has to decide whether (a) to allow the merger to go ahead unconditionally, (b) to allow it to go ahead subject to the firms meeting certain legally binding conditions, or (c) to prohibit the merger. In the 18 years between 2004/05 and 2021/22 only 186 of the 1587 cases considered were referred to Phase 2 and, of these, only 18 mergers were prohibited.

Pause for thought

If anti-monopoly legislation is effective enough, is there ever any need to prevent mergers from going ahead?

At the time of writing, there is some uncertainty about the long-term impact of the UK's exit from the European Union on competition policy. When the UK was a member of the EU, the system was referred to as a 'One Stop Shop'. Cases were investigated by either the CMA or the EC, but not by both of them. With the UK having left the Single Market, both the CMA and EC may have to investigate the same cases, leading to an increase in the workload of the CMA. Firms may also face a situation where two different agencies produce contradictory conclusions: i.e. one prohibits while the other clears the same merger.

UK regulation of privatised utilities

In the UK there are also regulatory bodies for each of the major privatised utilities. These regulators are as follows: the Office of Gas and Electricity Markets (Ofgem), the Office of Communications (Ofcom), the Office of Rail and Road (ORR) and the Water Services Regulation Authority (Ofwat).

As well as supervising the competitive behaviour of the privatised utility, they set terms under which the industries have to operate. For example, the ORR sets the terms under which rail companies have access to track and stations. The terms set by the regulator can be reviewed by negotiation between the regulator and the industry. If agreement cannot be reached, the CMA acts as an appeal court and its decision is binding.

The regulator for each industry also sets limits to the prices that certain parts of the industry can charge. These parts are those where there is little or no competition: for example, the charges made to electricity and gas retailers by National Grid, the owner of the electricity grid and major gas pipelines.

The price-setting formulae have largely been of the '*RPI* minus *X*' variety (although other factors, including competition and excessive profits are also taken into account). What this means is that the industries can raise their prices by the rate of increase in the retail price index (i.e. by the rate of inflation) *minus* a certain percentage (*X*) to take account of expected increases in efficiency. Thus if the rate of inflation were 3 per cent, and if the regulator considered that the industry (or firm) could be expected to reduce its costs by 2 per cent ($X = 2\%$), then price rises would be capped at 1 per cent.

The $RPI - X$ system is thus an example of **price-cap regulation**. The idea of this system of regulation is that it forces the industry to pass cost savings on to the consumer. (The $RPI - X$ system is considered further in Case Study 8.4 on the student website.)

In 2013, Ofgem introduced a new formula for price capping in the transmission and distribution networks in the gas and electricity markets. This is known as RIIO, where the firms' prices and hence permitted revenue (R) should not only depend on

Definition

Price cap regulation Where the regulator puts a ceiling on the amount by which a firm can raise its price.

costs, but should also have an element for incentives (I), innovation (I) and the quality of output (O). This means that firms' ability to raise prices is conditional on their performance in the following areas – levels of customer satisfaction, reliability, the conditions for connection, the environmental impact, social obligations and safety.

One challenge for the regulator is to find effective performance measures for each of these different attributes of the network companies' output.

Note that retail prices to consumers are largely not regulated as there is felt to be sufficient competition in this part of the market.

Pause for thought

What other forms of intervention are likely to be necessary to back up the work of regulatory bodies?

Recap

1. Laws can be used to regulate activities that impose external costs, to regulate monopolies and oligopolies, and to provide consumer protection. Legal controls are often simpler and easier to operate than taxes, and are safer when the danger is potentially great. Nevertheless, legal controls tend to be rather a blunt weapon, although discretion can sometimes be allowed in the administration of the law.
2. Regulatory bodies can be set up to monitor and control activities that might be against the public interest (e.g. anti-competitive behaviour of oligopolists).
3. Competition policy in most countries recognises that monopolies, mergers and restrictive practices can bring both costs and benefits to the consumer. Generally, though, restrictive practices tend to be more damaging to consumers' interests than simple monopoly power or mergers.
4. The focus of both EU and UK legislation is on anti-competitive practices rather than on the simple existence of agreements between firms or market dominance. Practices that are found after investigation to be detrimental to competition are prohibited and heavy fines can be imposed, even for a first offence.
5. Regulation in the UK of privatised utilities has involved setting up regulatory offices for the separate industries. These generally operate informally, using negotiation and bargaining to persuade the industries to behave in the public interest. They also set the terms under which the firms can operate (e.g. access rights to the respective grid).
6. As far as prices are concerned, in many cases privatised utilities are not controlled, as the industries are sufficiently competitive. Where the regulator deems that competition is insufficient, the industries are normally required to abide by an '*RPI* minus *X*' formula. This forces them to pass potential cost reductions on to the consumer.

8.7 OTHER FORMS OF GOVERNMENT INTERVENTION

What other means does the government have to correct market failures?

Changes in property rights

KI 28 p 196

One cause of market failure is the limited nature of property rights. If someone dumps a load of rubble in your garden, you can insist that it is removed. If, however, someone dumps a load of rubble in their *own* garden, but which is next door to yours, what can you do? You can still see it from your window. It is still an eyesore. But you have no property rights over the next-door garden.

Property rights define who owns property, to what uses it can be put, the rights other people have over it and how it may be transferred. By *extending* these rights, individuals may be able to prevent other people imposing costs on them or charge them for doing so.

The socially efficient level of charge would be one that was equal to the marginal external cost (and would have the same effect as the government charging a tax on the firm of that amount: see Figure 8.7). The **Coase theorem**[4] states that in an otherwise perfectly competitive market, the socially efficient charge *will* be levied. But why?

TC 12 p 195

Definition

The Coase theorem By sufferers from externalities doing deals with perpetrators (by levying charges or offering payments), the externality will be 'internalised' and the socially efficient level of output will be achieved (assuming zero bargaining costs).

[4]Named after Ronald Coase, who developed the theory. See his 'The problem of social cost', *Journal of Law and Economics* (1960).

KI 28 p 196 Let us take the case of river pollution by a chemical works that imposes a cost on people fishing in the river. If property rights to the river were now given to the fishing community, they could impose a charge on the chemical works per unit of output. If they charged *less* than the marginal external cost, they would suffer more from the last unit (in terms of lost fish) than they were being compensated. If they charged *more,* and thereby caused the firm to cut back its output below the socially efficient level, they would be sacrificing receiving charges that would be greater than the marginal suffering. It will be in the sufferers' best interests, therefore, to charge an amount *equal* to the marginal externality.

Pause for thought

If the sufferers had no property rights, show how it would still be in their interests to 'bribe' the firm to produce the socially efficient level of output.

In most instances, however, this type of solution is totally impractical. It is impractical when *many* people are *slightly* inconvenienced, especially if there are many culprits imposing the costs. For example, if I were disturbed by noisy lorries passing by my home, it would not be practical to negotiate with every haulage company involved. What if I wanted to ban the lorries from the street but my next-door neighbour wanted to charge them £1 per journey? Who gets their way?

The extension of private property rights becomes more practical where the parties involved are few in number, are easily identifiable and where the costs are clearly defined. Thus a noise abatement Act could be passed which allowed me to prevent my neighbours playing noisy radios, having noisy parties or otherwise disturbing the peace in my home. The onus would be on me to report them. Or I could agree not to report them if they paid me adequate compensation.

But even in cases where only a few people are involved, there may still be the problem of litigation. Justice may not be free, and there may be concerns that companies or rich individuals can afford 'better' justice than individuals on more modest incomes. The rich can employ top lawyers. Even if I have a right to sue a large company for dumping toxic waste near me, I may not have the legal muscle to win.

KI 24 p 185 Finally there is a broader question of *equity*. The extension of private property rights may favour the rich (who tend to have more property) at the expense of the poor. Ramblers may get great pleasure from strolling across a great country estate, along public rights of way. This may annoy the owner. If the owner's property rights are now extended to exclude the ramblers, is this a social gain?

Equity considerations can also be dealt with by altering property rights, but in a different way. *Public* property like parks, open spaces, libraries and historic buildings could be extended. Also the property of the rich could be redistributed to the poor. Here it is less a question of the rights that ownership confers, and more a question of altering the ownership itself.

Pause for thought

Would it be a good idea to extend countries' territorial waters in order to bring key open seas fishing grounds within countries' territory? Could this help to solve the problem of over-fishing?

Provision of information

When imperfect information is a reason for market failure, the direct provision of information by the government or one of its agencies may help to correct that failure. An example is the information on jobs provided by job centres to those looking for work. They thus help the labour market to work better and increase the elasticity of supply of labour.

Another example is the provision of consumer information – for example, on the effects of smoking, or of eating certain foodstuffs. Another is the provision of government statistics on prices, costs, employment, sales trends, etc. This enables firms to plan with greater certainty.

The direct provision of goods and services

In the case of public goods and services, such as streets, pavements, seaside illumination and national defence, the market may completely fail to provide. In this case the government must take over the role of provision. Central government, local government or some other public agency could provide these goods and services directly. Alternatively, they could pay private firms to do so. The public would pay through central and local taxation. KI 29 p 199

But just what quantity of the public good should be provided? How can the level of public demand or public 'need' be identified? Should any charge at all be made to consumers for each unit consumed?

With a pure public good, once it is provided the marginal cost of supplying one more consumer is zero. Take the case of a lighthouse used by passing pleasure craft. Once it is constructed and in operation, there is no extra cost of providing the service to additional passing craft. Even if it were *possible* to charge boats each time they make use of it, it would not be socially desirable. Assuming no external costs, *MSC* is zero. Thus $MSB = MSC$ at a price of zero. Zero is thus the socially efficient price. TC 12 p 195

But what about the construction of a new public good, like a new road or a new lighthouse? How can a rational decision be made by the government as to whether it should go ahead? This time the marginal cost is not zero: extra roads and lighthouses cost money to build. The solution is to identify all the costs and benefits to society from the project (private and external) and to weigh them up. This is known as **cost–benefit analysis** (see Web Appendix 8.2 for details of how cost–benefit analysis is conducted). If the social benefits of the project exceed the social costs, then it would be socially efficient to go ahead with it. Many proposed public projects are subjected to cost–benefit analysis in order to assess their desirability.

The government could also provide goods and services directly which are *not* public goods. Examples include health and education. We can identify various reasons why such things are provided free, or at well below cost, which link to one or more of the broad reasons for intervention identified in Section 1.5 (see page 26). These factors include:

KI 24 p 185 Social justice. Society may feel that these things should not be provided according to ability to pay. Rather they should be provided as of right: an equal right based on need.

KI 28 p 196 *Large positive externalities.* People other than the consumer may benefit substantially. If a person decides to get treatment for an infectious disease, other people benefit by not being infected. A free health service thus helps to combat the spread of disease.

Dependants. If education were not free, and if the quality of education depended on the amount spent, and if parents could choose how much or little to buy, then the quality of children's education would depend not just on their parents' income, but also on how much they cared. A government may choose to provide such things free in order to protect children KI 18 p 117 from 'bad' or 'foolish' parents. A similar argument is used for providing free prescriptions and dental treatment for all children.

Imperfect information. Consumers may not realise how much they will benefit. If they had to pay, they may choose (unwisely) to go without. Providing health care free may persuade people to consult their doctors before a complaint becomes serious.

Definition

Cost–benefit analysis The identification, measurement and weighing up of the costs and benefits of a project in order to decide whether or not it should go ahead.

Macroeconomic benefits. Public provision of services such as health and education can positively affect the economy's human capital, social mobility and social capital (see Box 7.5). In turn, these can positively impact on the effectiveness of the workforce and hence the economy's productive potential.

Nationalisation and privatisation

Another possible solution to market failure, advocated by some on the political left, is nationalisation. If industries are not being run in the public interest by the private sector, then bring them into public ownership. Problems of monopoly power, externalities, inequality, etc. can be dealt with directly if these industries are run with the public interest, rather than private gain, at heart.

In the late 1940s and early 1950s the Labour government of the time nationalised many of the key transport, communications and power industries, such as the railways, freight transport, airlines, coal, gas, electricity and steel.

However, by the mid-1970s the performance of the nationalised industries was being increasingly questioned. A change of policy was introduced in the early 1980s, when successive Conservative governments engaged in an extensive programme of 'privatisation', returning virtually all of the nationalised industries back to the private sector. These included telecommunications, gas, water, steel, electricity and the railways.

By 1997, the year the Conservatives left office, with the exception of the rail industry in Northern Ireland and the water industry in Northern Ireland and Scotland, the only nationalised industry remaining in the UK was the Post Office® (including post offices and mail). The Post Office and Royal Mail were split in 2012 and Royal Mail was privatised in October 2013. Post Office Ltd remains state owned but, under the 2011 Postal Services Act, there is the option for it to become a mutual organisation in the future.

Other countries have followed similar programmes of privatisation in what has become a worldwide phenomenon. Privatisation has been seen as a means of revitalising ailing industries and as an opportunity to raise revenues to ease budgetary problems.

In 2008, however, many governments returned to the use of nationalisation, in order to 'rescue' banks which were at risk of going bankrupt. This was intended to ensure the stability of the financial system – a theme we will return to in the *macro*economic chapters. In the EU, the process was facilitated by Member States being given permission to support financial institutions, subject to conditions under EU state aid rules.

In many countries, although some industries, or parts of industries, have been privatised, the process has not been as extensive as in the UK. In many countries, for example, the railways remain state owned.

In the UK there have been calls for renationalisation of certain privatised industries. In the 2019 general election, the Labour Party stated that, if elected, it would renationalise the train operating companies, the water industry, the energy industries, Royal Mail and fibre broadband provision – policies that were popular, according to opinion polls.

In response to criticisms of poor and expensive service on the railways, in 2021 the UK government announced that it would create of a new public body, Great British Railways (GBR), which would take over from Network Rail. GBR would develop a 30-year rail strategy and 5-year business plans, own and manage the infrastructure, set most fares and timetables, receive the fare revenue and procure passenger services. Private train operating companies would be paid fees to cover costs with incentives for efficiency.

Arguments for and against privatisation

The principal argument in favour of privatisation is that it exposes industries to *market forces*, which will stimulate greater efficiency, faster growth and greater responsiveness to consumers.

If an industry is split into competing companies, this greater competition in the goods market may force the companies to drive down costs and reduce prices in order to stay in business.

To finance investment, privatised companies must now go to the market: they must issue shares or borrow from banks or other financial institutions. In doing so, they will be competing for funds with other companies, and thus must be seen as capable of using these funds profitably.

Market discipline will also be enforced by shareholders. Shareholders want a good return on their shares and will thus put pressure on the privatised company to perform well. The competition for corporate control (see page 133) thus provides incentives for firms to be efficient.

Problems with privatisation. Some industries are a natural monopoly (see page 128). In such cases, the market forces argument for privatisation largely breaks down if a public monopoly is simply replaced by a private monopoly.

KI 19 p 122

Critics also point to the fact that various industries may create substantial external benefits and yet may be privately unprofitable. A railway or an underground line, for example, may considerably ease congestion on the roads, thus benefiting road as well as rail users. Other industries may cause substantial external costs. A coal-fired power station, for example, may emit large amounts of CO_2.

KI 28 p 196

For reasons of equity, it can be argued that various services, such as transport, be subsidised in order to keep them going and/or to keep their prices down. For example, it can be argued that rural bus services should be kept operating at subsidised prices and that certain needy people (e.g. pensioners) should be charged lower prices. Indeed, in 2007 the Labour government introduced an England-wide system of bus passes for people over 60, whereby bus journeys are free after the morning rush hour. The eligible age has since been raised to the state pension age (currently 66) but remains at 60 in London. The devolved administrations in the other parts of the UK have a similar system, with the eligible age being 60.

Generally, the solution to the problems of privatisation has been one of intervention (through taxes, subsidies, regulation, etc.) rather than renationalisation.

Recap

1. An extension of property rights may allow individuals to prevent others imposing costs on them, or to charge them for so doing. This is not practical, however, when many people are affected to a small degree, or where several people are affected but differ in their attitudes towards what they want done about the 'problem'.
2. The government may provide information in cases where the private sector fails to provide an adequate level.
3. The government may also provide goods and services directly. These could be in the category of public goods or other goods where the government feels that provision by the market is inadequate.
4. Another approach to market failure is to bring some industries under public control. However, many countries have over recent decades embarked on programmes of privatisation. One reason is to create competition, not only in the goods market but in the market for finance and for corporate control. However, critics point to problems of natural monopoly or significant externalities as well as to equity considerations. For this reason, there have been various forms of government intervention in privatised industries.

8.8 MORE OR LESS INTERVENTION?

Can the government always put things right?

Government intervention in the market can itself lead to problems. The advocates of privatisation, for example, argue that, as well as benefiting from the discipline of market forces, privatised industries are free of government interference and are able to make more rational economic decisions and plan with greater certainty.

However, the general case for less government intervention is not that the market is the *perfect* means of achieving given social goals, but rather that the problems created by intervention are greater than the problems overcome by that intervention.

Drawbacks of government intervention

TC 6 p 24

Lack of market incentives. Though the market may be imperfect, it does tend to encourage efficiency by allowing the efficient to receive greater rewards. Government intervention, whether this is nationalisation, welfare payments, the use of subsidies, guaranteed prices or wages etc., removes market forces or cushions their effect. Therefore, it may remove certain useful incentives. Subsidies, for example, may allow inefficient firms to survive, while welfare payments may be argued to discourage effort.

Shifts in government policy. The economic efficiency of industry may suffer if government intervention changes too frequently. It makes it difficult for firms to plan if they cannot predict tax rates, subsidies, price and wage controls, etc.

KI 8 p 24

Shortages and surpluses. If the government intervenes by fixing prices at levels other than the equilibrium, this will create either shortages or surpluses (see Section 3.6).

If the price is fixed *below* the equilibrium, there will be a shortage. For example, if the rent of social housing is fixed below the equilibrium in order to provide affordable housing for low-income households, demand will exceed supply. In the case of such shortages the government will have to adopt a system of waiting lists, or rationing, or giving certain people preferential treatment. Alternatively it will have to allow allocation to be on a first-come, first-served basis or allow queues to develop. Underground (or 'shadow') markets are likely to occur (see page 66).

If the price is fixed *above* the equilibrium price, there will be a surplus. Such surpluses are obviously wasteful.

Poor information. The government may not know the full costs and benefits of its policies. It may genuinely wish to pursue the interests of consumers or any other group and yet may be unaware of people's wishes or misinterpret their behaviour.

Bureaucracy and inefficiency. Government intervention involves administrative costs. The more wide-reaching and detailed the intervention, the greater the number of people and material resources that will be involved. These resources may be used wastefully.

Lack of freedom for the individual. Government intervention involves a loss of freedom for individuals to make economic choices. The argument is not just that the pursuit of individual gain is seen to lead to the social good, but that it is desirable in itself that individuals should be as free as possible to pursue their own interests with the minimum of government interference: that minimum being largely confined to the maintenance of laws consistent with the protection of life, liberty and property.

Advantages of the free market

Although markets in the real world are not perfect, even imperfect markets can be argued to have positive advantages over government provision or even government regulation. These might include the following:

KI 8 p 24

Automatic adjustments. Government intervention requires administration. A free-market economy, on the other hand, leads to automatic, albeit imperfect, adjustment to demand and supply changes.

TC 6 p 24

Dynamic advantages of capitalism. The chances of making high monopoly/oligopoly profits will encourage entrepreneurs to invest in new products and new processes and techniques. It thereby encourages innovation which has macroeconomic benefits too. While prices may be high initially, consumers will gain from the extra choice of products. Furthermore, if profits are high, new firms will sooner or later break into the market and competition will ensue.

Pause for thought

Are there any features of free-market capitalism that would discourage innovation?

A high degree of competition even under monopoly/oligopoly. Even though an industry at first sight may seem to be highly monopolistic, competitive forces may still work as a result of the following:

- A fear that excessively high profits might encourage firms to attempt to break into the industry (assuming that the market is contestable).
- Competition from closely related industries (e.g. coach services for rail services, or electricity for gas).
- The threat of foreign competition.
- Countervailing powers (see page 147). Large powerful producers often sell to large powerful buyers. For example, the power of detergent manufacturers to drive up the price of washing powder is countered by the power of supermarket chains to drive down the price at which they purchase it. Thus power is to some extent neutralised.
- The competition for corporate control (see page 133).

Should there be more or less intervention in the market?

No firm conclusions can be drawn in the debate between those who favour more and those who favour less government intervention, for the following reasons:

- The debate involves many moral, social and political issues which cannot be settled by economic analysis. For example, it could be argued that freedom to set up in business and freedom from government regulation are desirable for their own sake. As a fundamental ethical point of view this can be disputed, but not disproved.
- In principle, the issue of whether a government ought to intervene in any situation could be settled by weighing up the costs and benefits of that intervention. However, such costs and benefits, even if they could be identified, are extremely difficult if not impossible to measure, especially when the costs are borne by different people from those who receive the benefits and when externalities are involved.
- Often the effect of more or less intervention simply cannot be predicted: there are too many uncertainties.

Nevertheless, economists can make a considerable contribution to analysing problems of the market and the effects of government intervention.

Recap

1. Government intervention in the market may lead to shortages or surpluses; it may be based on poor information; it may be costly in terms of administration; it may stifle incentives; it may be disruptive if government policies change too frequently; it may remove certain liberties.
2. By contrast, a free market leads to automatic adjustments to changes in economic conditions; the prospect of monopoly/oligopoly profits may stimulate risk-taking and hence research and development and innovation; there may still be a high degree of actual or potential competition under monopoly and oligopoly.
3. It is impossible to draw firm conclusions about the 'optimum' level of government intervention. This is partly due to the moral/political nature of the question, partly due to the difficulties of measuring costs and benefits of intervention/non-intervention, and partly due to the difficulties of predicting the effects of government policies, especially over the longer term.

8.9 THE ENVIRONMENT: A CASE STUDY IN MARKET FAILURE

How can economists contribute to the environmental debate?

The environmental problem

Scarcely a day goes by without some environmental issue or other featuring in the news: another warning about global warming; a company fined for illegally dumping waste; a drought or flood blamed on pollution/global warming; smog in some major cities. Also attempts by policy makers to improve the environment are often controversial and hit the headlines: for example, the impact of government climate change policies on the size of people's energy bills.

Why does the environment appear to be so misused and policies that attempt to improve the situation so controversial? To answer these questions we have to understand the nature of the economic relationship between humans and the natural world.

The environment as a resource

We all benefit from the environment in three ways:

- as an amenity to be enjoyed
- as a source of primary products (food, raw materials and other resources)
- as a place where we can dump waste.

Unfortunately, these three different uses are often in conflict with each other. For example, we extract and burn fossil fuels such as coal, oil and gas, for power generation and industrial uses. However, the extraction of these fuels may have a negative impact on the amenity value of the environment. One only has to think of some of the concerns raised by people about the impact of drilling for shale gas on the appearance of their local communities. The burning of fossil fuels also creates greenhouse gases that are emitted into the atmosphere and cause climate change. Some of the CO_2 gets absorbed into the oceans, which increases their level of acidity and kills marine life.

KI 28 p 196

Policies that try to reduce our current use of the environment as a source of primary products and/or reduce the volume of emissions we generate come at cost. These higher costs are often passed on to consumers in the form of higher prices – an outcome they often dislike and complain about.

The subject of environmental degradation lies clearly within the realm of economics, since it is a direct consequence of production and consumption decisions. So how can economic analysis help us to understand the nature of the problem and design effective policies that will result in the optimal use of the environment? What will be the impact of these policies on households and firms?

Market failures

KI 25 p 195

An unregulated market system may fail to provide an adequate protection for the environment for a number of reasons.

KI 28 p 196

Externalities. We saw in Section 8.2 how pollution could be classified as a 'negative externality' of production or consumption. In the case of production, there are marginal external costs (*MEC*), which means that the marginal social costs (*MSC*) are greater than the marginal private costs (*MC*) to the polluter.

The failure of the market system to equate *MSC* and marginal social benefit (*MSB*) is due to either consumers or firms lacking the appropriate property rights. Because no one owns the environment, there is no one to enforce property rights over it. If a company pollutes the air that I breathe, I cannot stop it because the air does not belong to me.

KI 29 p 199

The environment as a common resource. The air, the seas and many other parts of the environment are not privately owned. They are a global 'commons'. As such, it is extremely difficult to exclude non-payers from consuming the benefits they provide. Because of this property of 'non-excludability' (see page 199), the environment can often be consumed at a zero price. If the price of any good or service to the user is zero, there is no incentive to economise on its use.

Many parts of the environment, however, are *scarce*: there is *rivalry* in their use. As people increase their use of the environment, it may prevent other or rival consumers from enjoying it. We saw in Box 8.1 how over-fishing in the open oceans can lead to the depletion of fish stocks.

Ignorance. There have been many cases of people causing environmental damage without realising it, especially when the effects build up over a long time. Take the case of aerosols. It was not until the 1980s that scientists connected their use to ozone depletion. Even when the problems are known to scientists, consumers may not appreciate the full environmental costs of their actions. So even if people would like to be more 'environmentally friendly' in their activities, they might not have the knowledge to be so.

Inter-generational problems. The environmentally harmful effects of many activities are long term, whereas the benefits are immediate. Thus consumers and firms are frequently prepared to continue with various practices and leave future generations to worry about their environmental consequences. The problem, then, is a reflection of the importance that people attach to the present relative to the future.

KI 30 p 203

Pause for thought

Look through the categories of possible market failings in Sections 8.1 to 8.4. Are there any others, in addition to the four we have just identified, that will result in a socially inefficient use of the environment?

Policy alternatives

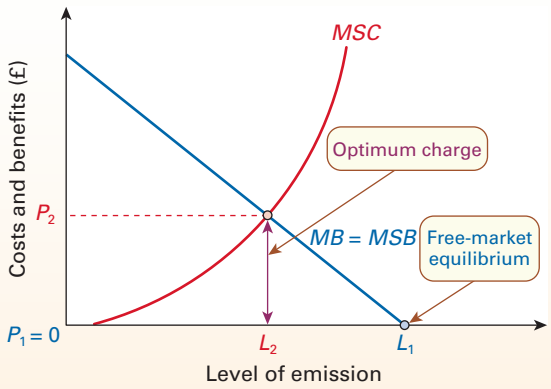
Charging for use of the environment (as a resource or a dump)

One way of 'pricing the environment' is for the government to impose **environmental charges** on

Definition

Environmental charges Charges for using natural resources (e.g. water or national parks), or for using the environment as a dump for waste (e.g. factory emissions or sewage).

Figure 8.9 An emissions charge



consumers or firms. Thus *emissions charges* could be levied on firms discharging waste. Another example is the use of *user charges* to households for sewage disposal or rubbish collection. The socially efficient level of environmental use would be where the marginal social benefits and costs of that use were equal. This is illustrated in Figure 8.9, which shows the emission of toxic waste into a river by a chemical plant.

It is assumed that all the benefits from emitting the waste into the river accrue to the firm (i.e. there is no external benefit). Marginal private and marginal social benefits are thus the same ($MB = MSB$). The curve slopes downwards because, with a downward-sloping demand curve for the *good*, higher output will have a lower marginal benefit, and so too will the waste associated with it.

But what about the marginal costs? Without charges, the marginal private cost of using the river for emitting the waste is zero. The pollution of the river, however, imposes an external cost on those living by the river or using it for fishing or water supply. The marginal external cost *rises* as the river becomes less and less able to cope with increased levels of effluent. As there is no private cost, the marginal social cost is the same as the marginal external cost.

Without a charge, the firm will emit L_1, since this is where its private marginal cost ($= 0$) equals its private marginal benefit. The socially efficient level of emission is L_2 and the socially efficient level of emission charge, therefore, is P_2.

Environmental ('green') taxes and subsidies

Rather than charging for environmental use, a tax could be imposed on the output (or consumption) of a good, wherever external environmental costs are generated. These are known as **green taxes**. In this case, the good already has a price: the tax has the effect of increasing the price. To achieve a socially efficient output, the rate of tax should be equal to the marginal external cost (see Figure 8.7 on page 207). The alternative is to subsidise activities that reduce pollution (such as the installation of loft insulation). Here the rate of subsidy should be equal to the marginal external benefit.

Although green taxes and subsidies are theoretically a means of achieving social efficiency, they do have serious limitations (see Box 8.3).

Laws and regulations

The traditional way of tackling pollution has been to set maximum permitted levels of emission or resource use, or minimum acceptable levels of environmental quality, and then to fine firms contravening these limits. Measures of this type are known as **command-and-control (CAC) systems**. Clearly, there have to be inspectors to monitor the amount of pollution, and the fines have to be large enough to deter firms from exceeding the limit.

Virtually all countries have environmental regulations of one sort or another. For example, the EU has over 230 items of legislation covering areas such as air and water pollution, noise, the marketing and use of dangerous chemicals, waste management, the environmental impacts of new projects (such as power stations, roads and quarries), recycling, depletion of the ozone layer and global warming.

Given the uncertainty over the environmental impacts of pollutants, especially in the longer term, it is often better to play safe and set tough emissions standards or standards for environmental impact ('ambient' standards). These could always be relaxed at a later stage if the effects turn out not to be so damaging, but it might be too late to reverse damage if the effects turn out to be more serious. Taxes may be a more sophisticated means of reaching a socially efficient output, but regulations are usually more straightforward to devise, easier to understand by firms and easier to implement.

Education

People's attitudes are very important in determining the environmental consequences of their actions. Fortunately for the environment, people are not always out simply to maximise their own self-interest.

Definitions

Green tax A tax on output designed to charge for the adverse effects of production on the environment. The socially efficient level of a green tax is equal to the marginal environmental cost of production.

Command-and-control (CAC) systems The use of laws or regulations backed up by inspections and penalties (such as fines) for non-compliance.

BOX 8.3 GREEN TAXES

Are they the answer to the problem of pollution?

Countries are increasingly using 'green' taxes in order to discourage pollution as goods are produced, consumed or disposed of. The Office for National Statistics estimated that environmental taxes raised £52.620 billion of revenue for the UK government in 2019 (7.1 per cent of total taxation and social security receipts), while households paid an average of £760 on green taxes in 2018 – 35 per cent more than in 1997.

The table shows some of the more important environmental taxes that have been used in the UK. Taxes on hydrocarbon oils (i.e. duty on petrol and diesel) are by far the most significant and account for nearly 60 per cent of total revenue from all environmental taxes. Note that the figures are for 2019 as 2020 and 2021 figures were influenced by the pandemic.

Government revenue from environmental taxes, 2021

		£m	% of environmental taxes
Energy taxes	Tax on hydrocarbon oils	27 795	52.8
	Climate change levy	2091	4.0
	Renewable energy obligations	6118	11.6
	EU Emissions Trading Scheme	1255	2.4
	Other	1769	3.4
Transport taxes	Air Passenger Duty	3810	7.2
	Rail Franchise Premia	1258	2.4
	Motor vehicle duties paid by businesses	2007	3.8
	Motor vehicle duty paid by households	4998	9.5
	Other	312	0.6
Pollution and resources taxes	Landfill tax	784	1.5
	Aggregates levy	396	0.8
	Fishing licences	27	0.1
Total		52 620	100.0

Source: Based on data from *Environmental taxes in the United Kingdom* (Office for National Statistics)

Environmental tax revenue in the UK as a percentage of GDP has remained broadly stable since 2000 at between 2¼ and 2¾ per cent. The chart shows how the UK compares with other countries.

As can be seen, green taxes are typically higher than average in Scandinavian and some other European countries, reflecting the strength of their environmental concerns. They are lowest in the USA. By far the largest green tax revenues come from fuel taxes. Fuel taxes are relatively high in the UK and so, therefore, are green tax revenues.

There are various problems, however, with using taxes in the fight against pollution.

Identifying the socially efficient tax rate. It will be difficult to identify the appropriate amount of tax for each firm, given that each one is likely to produce different amounts of pollutants for any given level of output. Even if two firms produce identical amounts of pollutants, the environmental damage might be quite different, because the ability of the environment to cope with it will differ between the two locations. Also, the number of people suffering will differ (a factor that is very important when considering the *human* impact of pollution). What is more, the harmful effects are likely to build up over time, and predicting these effects is fraught with difficulty.

Problems of demand inelasticity. The less elastic the demand for the product, the less effective will a tax be in cutting production and hence in cutting pollution. Thus taxes on petrol would have to be very high indeed to make significant reductions in the consumption of petrol and hence significant reductions in the exhaust gases that contribute towards global warming and acid rain.

KI 12 p 49

Problems with international trade. If a country imposes pollution taxes on its industries, its products will become less competitive in world trade. To compensate for this, it may be necessary to give the industries tax rebates for exports. Also, taxes would have to be imposed on imports of competitors' products from countries where there is no equivalent green tax.

Evidence on the adverse effect of environmental taxes on a country's exports is inconclusive, however. Over the long term, in countries with high environmental taxes (or other tough environmental measures), firms will be stimulated to invest in low-pollution processes and products. This will later give such countries a competitive advantage if *other* countries then impose tougher environmental standards.

For example, they are sometimes willing to pay higher prices for 'green products'. There is evidence that people's willingness to pay has increased markedly over the past few decades and there is more pressure on firms to improve their 'green credentials'.

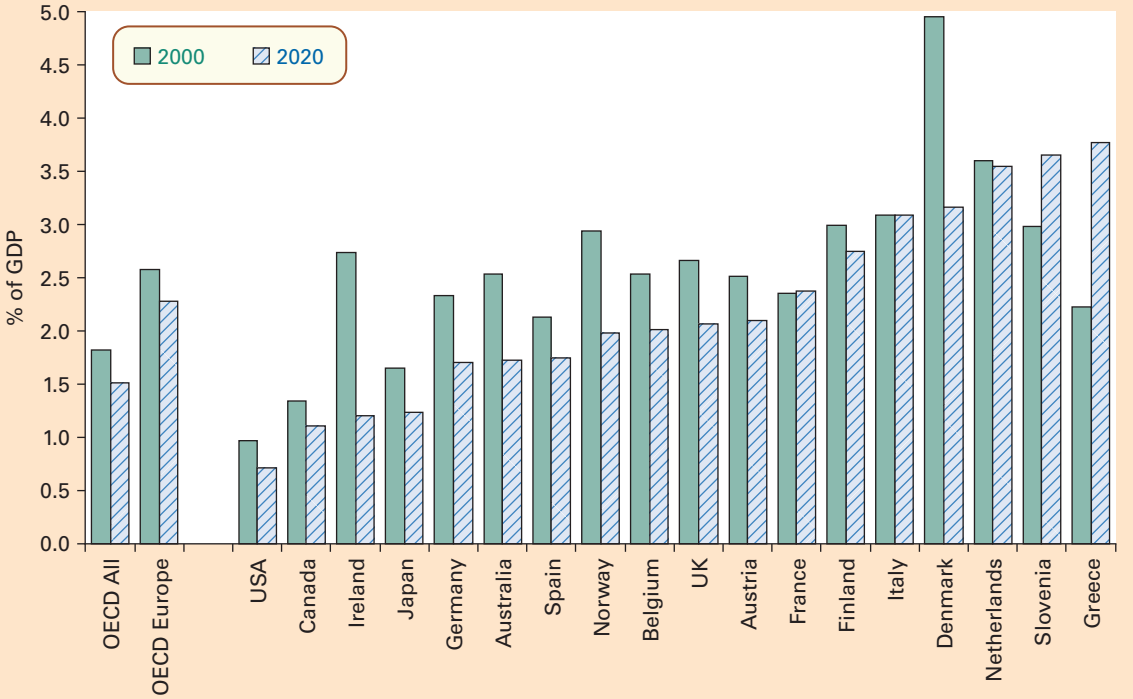
Education can play a significant part here. It can make people more aware of environmental issues and the consequences of their actions. It can thereby help to influence consumer behaviour and the pressure this puts on companies.

Tradable permits

A policy measure that has grown in popularity in recent years is that of **tradable permits** – sometimes known as **cap and trade.** This is a combination of regulations (CAC) and market-based systems.

CASE STUDIES & APPLICATIONS

Green tax revenues as a percentage of GDP



Note: Figures for Canada are for 2000 and 2014, for the USA are for 2000 and 2016 and for Australia are for 2000 and 2018

Source: Based on data in *Environmentally Related Taxes Database* (OECD, 2022)

Effects on employment. Reduced output in the industries affected by green taxes will lead to a reduction in employment. If, however, the effect was to encourage investment in new cleaner technology, employment might not fall. Furthermore, employment opportunities could be generated elsewhere, if the extra revenues from the green taxes were spent on alternative products (e.g. buses and trains rather than cars).

Redistributive effects. Many green taxes are regressive. The poor spend a higher proportion of their income on domestic fuel than the rich. A 'carbon tax' on such fuel therefore has the effect of redistributing incomes away from the poor. The poor also spend a larger proportion of their income on food than do the rich. Taxes on agriculture, designed to reduce the intensive use of fertilisers and pesticides, also tend to hit the poor proportionately more than the rich.

Not all green taxes, however, are regressive. The rich spend a higher proportion of their income on motoring than the poor. Thus petrol and other motoring taxes could have a progressive effect.

Despite these problems, such taxes can still move output closer to the socially efficient level. What is more, they do have the major advantage of providing a continuing incentive to firms to find cleaner methods of production and thereby save more on their tax bills.

Is it a good idea to use the revenues from green taxes to subsidise green alternatives (e.g. using petrol taxes for subsidising rail transport)? Consider the implications for wider tax policy in your answer.

Using data from Environmental taxes in the United Kingdom (Office for National Statistics) construct a pie chart for the latest year's data to show government revenue from environmental taxes split by economic activity (industrial type).

Definition

Tradable permits (**cap and trade**) Each firm is given a permit to produce a given level of pollution. If less than the permitted amount is produced, the firm is given a credit. This can then be sold to another firm, allowing it to exceed its original limit.

Capping pollution (a form of regulation). Initially some criteria have to be set in order to determine which factories, power plants and installations will be covered by the scheme. Policy makers then have to set a limit or 'cap' on the total volume of pollution these organisations will be collectively allowed to produce before any financial penalties are incurred.

BOX 8.4 TRADING OUR WAY OUT OF CLIMATE CHANGE

The EU carbon trading system

The EU introduced a carbon Emissions Trading Scheme (EU ETS) in January 2005 as its principal policy to meet environmental targets set by the international treaty, the Kyoto Protocol (which entered into force in February 2005). Article 17 of this treaty supported the use of emissions trading and a similar scheme had already reduced emissions of both sulphur dioxide and nitrous oxide in the USA. The EU ETS created a market in carbon permits or allowances.

Phases I and II

The first phase of the scheme ran from January 2005 until December 2007. Around 12 000 industrial plants across 27 countries were allocated approximately 2.2 billion CO_2 permits, called Emission Unit Allowances (EUAs). Each EUA issued to a firm gives it the right to emit one tonne of carbon dioxide into the atmosphere. The factories covered by the scheme were collectively responsible for around 40 per cent of the EU's CO_2 emissions each year.

Companies that do not have enough EUAs to match their annual emissions can purchase additional EUAs to cover the difference, while those that reduce their emissions are able to sell any surplus EUAs for a profit. Companies are able to trade directly with each other or via brokers operating throughout Europe.

At the end of December 2007 all existing allowances became invalid and the second Trading Period began, to last until the end of 2012. Although this was run under the same general principles as Trading Period 1, it also allowed companies to use 'Joint Implementation' and 'Clean Development Mechanism' credits earned under the Kyoto Protocol's project-based mechanisms (see Case Study 8.8 on the student website). In other words, companies could offset emissions *in* the EU against emission reductions they achieve in countries *outside* the EU.

Phase III

Phase III of the EU ETS came into operation on 1 January 2013. It built on the experience gained from operating Phases I and II of the system and included two significant changes.

Move to an EU-wide cap. The cap on total emissions in both Phases I and II of the system were set in a decentralised manner. Each member state had to develop a National Allocation Plan (NAP). The NAP set out the total cap on emissions for that country, the total quantity of EUAs that would be issued and how they would be assigned to each industrial plant or factory. Each NAP had to be approved by the European Commission before it could be implemented. The numerous NAPs have been replaced in Phase III of the EU ETS by a single EU-wide cap on the volume of emissions and on the total number of EUAs to be issued. The size of this EU-wide cap is to be reduced by 1.74 per cent per year so that emissions in 2020 are 21 per cent lower than 2005.

Move to auctioning permits. In Phases I and II of the EU ETS the majority of EUAs were freely allocated to the plants and factories covered by the scheme. The number of EUAs each factory would receive was based on their current emissions. This method is known as ***grandfathering***. The European Commission allowed member states to auction up to a maximum of 5 per cent of the EUAs in Phase 1 and 10 per cent in Phase II. However, this option was seldom chosen.

Phase III saw a big increase in the proportion of EUAs auctioned. From 2013 most of the firms in the power sector had to purchase all of their allowances by auction, while the average in other sectors was expected to increase from 20 per cent in 2013 to 70 per cent by 2020. Only firms in manufacturing and the power industry in certain member states would continue to be allocated the majority of their allowances at no charge.

It was also recommended by the EU that half of the revenue generated from the auctions should be used to fund measures to reduce greenhouse gas emissions.

Phase IV

Proposals agreed in October 2014 and revised in 2021 saw further revisions to the EU ETS for the period from 2021 to

Once an aggregate cap has been set, pollution permits (or 'allowances') for that total amount are either issued or sold to the firms. An alternative, and increasingly common, approach, is to auction the allowances. Each allowance held by a firm gives it the right to produce a given volume or weight of pollutants.

The total quantity of allowances awarded in subsequent years is then reduced by a certain percentage – the cap is tightened – to give firms an incentive to invest in cleaner technology.

All firms covered by the scheme must monitor and report the level of pollutants caused by their production activity. At the end of the year they must then submit enough allowances to the authorities to match the level of pollutants they have emitted. Each allowance can only be used once. If a firm fails to submit enough allowances, then it is subject to heavy fines.

Trading under a cap-and-trade system. Firms are able to buy and sell allowances in a secondary market once they have been allocated by the authorities.

Take the example of an organisation that estimates it will not have enough permits at the end of the year to match its forecast level of emissions. It can buy allowances in the secondary market from firms who have excess permits and hence are willing to sell.

The price that firms pay either to buy or to sell the allowances in the secondary market will depend

EXPLORING ECONOMICS

2030. These are designed to reduce the EU's greenhouse gas emissions by at least 55 per cent from 1990s' levels by 2030. The proposals are outlined in the EU's 2030 climate & energy framework as part of its contribution to the UN Paris Agreement on climate change. To achieve this the overall number of emission allowances will fall at an annual rate of 2.2 per cent from 2021 (rather than 1.74 per cent in Phase III). Also, 24 per cent of allowances would be 'set aside' in a Market Stability Reserve (MSR). All allowances in the MSR above the total number of allowances auctioned during the previous year would become invalid. The hope is that, by withdrawing allowances, the carbon price will rise.

In addition to 55 per cent cuts in greenhouse gas emissions, the 2030 climate & energy framework contains two further targets: at least a 32 per cent share for renewable energy and at least a 32.5 per cent improvement in energy efficiency.

Assessing the ETS

It remains difficult to assess the impact of the EU ETS, despite having been in operation since 2005. Disaggregating the effect of emission allowances from the effects of other economic factors and policy changes is enormously complicated.

What matters crucially for the impact of the scheme is the size of the overall cap and hence the total number of permits issued by the authorities. If the supply of the permits exceeds demand in the secondary market then the price will be relatively low and firms will lack the necessary incentives to invest in new energy efficient technology.

Some people have argued that the number of Emission Unit Allowances issued in the past has been far too generous. This problem was particularly acute in Phase I of the scheme. Emission levels across the EU rose by 1.9 per cent while the price of EUAs fell from a peak of €30 to just €0.02. Continuing surpluses were then fuelled by the economic and financial crisis of the late 2000s which reduced emissions as activity levels fell.

The excess supply of allowances remains an issue. The EU has tried to address this problem by introducing *backloading*. This involves withholding allowances from the auction process until later in the phases but without affecting the overall allowances issued. In Phase III auction volumes were reduced by 400 million in 2014, 300 million in 2015 and 200 million in 2016.

Managing any future supply—demand imbalances continues to be a major priority for EU policy makers. Hence, the process of backloading is continuing in Phase IV.

With the UK's exit from the EU, it opted for a standalone UK ETS scheme which was launched on 19 May 2021. As with the EU ETS, it is limited to energy-intensive industries (such as steelmaking), power generation and aviation. With some exceptions, allowances are auctioned, with a reserve price of £22 per tonne.

Consider a situation where all firms are of identical size and each is allocated credits that allows it to produce 10 per cent less than its current emissions. How would this compare with a situation where permits are allocated to 90 per cent of firms only? Consider both efficiency and equity in your answer.

Undertake desktop research to write a short briefing note on the climate strategies and targets of the European Union.

Definition

Grandfathering Where the number of emission permits allocated to a firm is based on its current levels of emission (e.g. permitted levels for all firms could be 80 per cent of their current levels).

on the levels of demand and supply. These levels will be heavily influenced by the initial number of permits allocated by the authorities, the state of the economy and developments in technology.

The principle of tradable permits can be used as the basis of international agreements on pollution reduction. Each country could be required to achieve a certain percentage reduction in a pollutant (e.g. carbon dioxide, CO_2, or sulphur dioxide, SO_2), but any country exceeding its reduction could sell its right to these emissions to other (presumably richer) countries. There are 24 emissions trading schemes (as of 2022) operating across the world at international, national, regional and city level (see Box 8.4 for an examination of the largest of these – the EU Emissions Trading Scheme (EU ETS)).

A similar principle can be adopted for using natural resources. Thus fish quotas could be assigned to fishing boats, or fleets or countries. Any parts of these quotas not used could then be sold.

Assessing the system of tradable permits. The main advantage of tradable permits is that they combine the simplicity of regulations with the benefits of achieving pollution reduction in the most efficient way. There is also the advantage that firms have a financial incentive to cut pollution. This might then make it easier for governments to impose tougher standards (i.e. impose lower permitted levels of emission).

TC 6 p 24

There are, however, various potential problems with tradable permits. One is the possibility that trade will lead to pollution being concentrated in certain

BOX 8.5 THE PROBLEM OF URBAN TRAFFIC CONGESTION

Does Singapore have the answer?

Singapore has some 283 vehicles per kilometre of road (this compares with 271 in Hong Kong, 222 in Japan, 77 in the UK, 75 in Germany and 37 in the USA). The average car in Singapore is driven some 17 500 kilometres per year, but, with low car ownership (see below), this translates into a relatively low figure for kilometres travelled by car per person. What is more, cars flow relatively freely: the average car speed during peak hours is estimated to be as high as 29 km/h on main roads and 64 km/h on expressways. Are there lessons that we can therefore learn from Singapore in tackling traffic congestion?

The problem of traffic congestion

KI 28 p 196

Traffic congestion is a classic example of the problem of externalities. When people use their cars, not only do they incur private costs (petrol, wear and tear on the vehicle, tolls, the time taken to travel, etc.), but also they impose costs on other people. These external costs include the following:

Congestion costs: time. When a person uses a car on a congested road, it will add to the congestion. This will therefore slow down the traffic even more and increase the journey time of other car users.

Congestion costs: monetary. Congestion increases fuel consumption, and the stopping and starting increases the costs of wear and tear. When a motorist adds to congestion, therefore, there will be additional monetary costs imposed on other motorists.

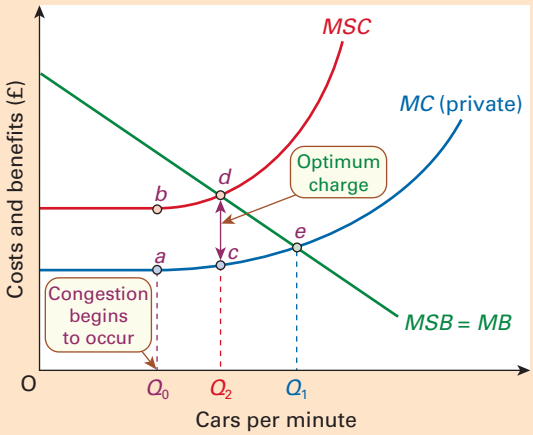
Environmental costs. When motorists use a road they reduce the quality of the environment for others. Cars emit fumes and create noise. This is bad enough for pedestrians and other car users, but can be particularly distressing for people living along the road. Driving can cause accidents – a problem that increases as drivers become more impatient because of delays.

Exhaust gases cause long-term environmental damage and are one of the main causes of the greenhouse effect and of the increased acidity of lakes and rivers and the poisoning of forests. They can also cause long-term health problems (e.g. for asthma sufferers).

The socially efficient level of road usage

These externalities mean that road usage will be above the social optimum. This is illustrated in the diagram. Costs and benefits are shown on the vertical axis and are measured in money terms. Thus any non-monetary costs or benefits (such as time costs) must be given a monetary value. The horizontal axis measures road usage in terms of cars per minute passing a specified point on the road.

Actual and optimum road usage



For simplicity it is assumed that there are no external benefits from car use and that therefore marginal private and marginal social benefits are the same. The *MSB* curve is shown as downward sloping. The reason for this is that different road users put a different value on any given journey. If the marginal (private) cost of making the journey were high, only those for whom the journey had a high marginal benefit would travel along the road. If the marginal cost of making the journey fell, more people would make the journey: people choosing to make the journey as long as the marginal cost of the journey was less than the marginal benefit. Thus the greater the number of cars, the lower the marginal benefit.

The marginal (private) cost curve (*MC*) is likely to be constant up to the level of traffic flow at which congestion begins to occur. This is shown as point a in the diagram. Beyond this point, marginal cost is likely to rise as time costs increase (i.e. journey times lengthen) and as fuel consumption rises.

The marginal social cost curve (*MSC*) is drawn above the marginal private cost curve. The vertical difference between the two represents the external costs. Up to point b, external costs are simply the environmental costs. Beyond point b, there are also external congestion costs, since additional road users slow down the journey of other road users. These external costs get progressively greater as traffic grinds to a halt.

The actual level of traffic flow will be at Q_1 where marginal private costs and benefits are equal (point e). The socially efficient level of traffic flow, however, will be at the lower level of Q_2 where marginal social costs and benefits are equal (point d). In other words, there will be an excessive level of road usage.

TC 12 p 195

So what can governments do to 'internalise' these externalities?

The Singapore solution

In contrast to its neighbours, many of which are suffering more acute urban traffic congestion problems, Singapore has an integrated transport policy. This includes restricting the number of new car licences and allowing their price to rise to the corresponding equilibrium. This makes car licences in Singapore among the most expensive in the world.

In addition, Singapore has a 200 kilometre-long Mass Rapid Transit (MRT) system with subsidised fares, a programme of building new estates near MRT stations and cheap, frequent buses, serving all parts of the island.

But it is in respect to road usage that the Singaporean authorities have been most innovative.

Area licences. The first innovation came in 1975 when the Area Licensing Scheme (ALS) was introduced. The city centre was made a restricted zone. Motorists who wished to enter this zone had to buy a ticket (an 'area licence') at any one of 33 entry points. Police were stationed at these entry points to check that cars had paid and displayed. This scheme was extended to the major expressways in 1995 with the introduction of the Road Pricing Scheme (RPS).

The Vehicle Quota System. Then in 1990 a quota system for new cars was established. The government decides the total number of cars that the country should have, and issues just enough licences each month to maintain that total. These licences (or 'Certificates of Entitlement (COEs)') are for 10 years and are offered at monthly auctions. These are operated in a similar manner to those on the eBay system. Buyers specify a maximum price and bids are automatically revised upwards until that maximum price is reached. The price of COEs increases until the quantity demanded is just equal to the number of certificates on offer.

Partly as a result of the quota system, there are only 98 private cars per 1000 population. This is only a fraction of the figure for European countries.

A problem with the licences is that they are a once-and-for-all payment, which does not vary with the amount that people use their car. In other words, their marginal cost (for additional miles driven) is zero. Many people feel that, having paid such a high price for their licence, they ought to use their car as much as possible in order to get value for money!

Electronic road pricing. With traffic congestion steadily worsening, it was recognised that something more had to be done. Either the Area Licensing Scheme had to be widened, or some other form of charging had to be adopted. The decision was taken to introduce Electronic Road Pricing Scheme (ERP). This alternative would not only save on police labour costs, but enable charge rates to be varied according to levels of congestion, times of the day, and locality.

What, then, would be the optimum charge? If the objective is to reduce traffic from Q_1 to Q_2 in the diagram, then a charge of $d - c$ should be levied.

All vehicles in Singapore are fitted with an in-vehicle unit (IU). Every journey made requires the driver to insert a smart card containing pre-paid units into the IU. On specified roads, overhead gantries read the IU and deduct the appropriate charge from the card. If a car does not have sufficient funds on its smart card, the car's details are relayed to a control centre and a fine is imposed. The system has the benefit of operating on three-lane highways and does not require traffic to slow down.

The ERP system operates on roads subject to congestion and charges can vary every 5, 20 or 30 minutes according to predicted traffic flows. Rates are published in advance for a three-month period. A review of traffic conditions takes place every quarter and the results can lead to rates being adjusted in future periods. The system is thus very flexible to allow traffic to be kept at the desired level.

Satellite tracking. The authorities in Singapore are now installing a Global Navigation Satellite System. The original date for completion was 2021 but because of the global shortage of computer chips this was pushed back to the second half of 2023. This new system removes the need for large overhead gantries, although slimmer gantries will still be used to indicate the charges.

The ERP system was expensive to set up. Cheaper schemes have been adopted elsewhere, such as Norway and in parts of the USA. These operate by funnelling traffic into a single lane in order to register the car, but they have the disadvantage of slowing the traffic down.

One message is clear from the Singapore solution. Road pricing alone is not enough. Unless there are fast, comfortable and affordable public transport alternatives, the demand for cars will be highly price inelastic. People have to get to work!

Explain how, by varying the charge debited from the smart card according to the time of day or level of congestion, a socially optimal level of road use can be achieved.

Research the alternative ways in which various towns or cities have attempted to deal with their traffic congestion problems. Summarise these approaches and discuss the underlying economic principles.

geographical areas. Another is that it may allow dirtier factories (or countries) to continue with high emissions if they can simply buy permits. Additionally, the system will lead to significant cuts in pollution only if the permitted levels are low. Once the system is in place, the government might then feel that the pressure is off to *reduce* the permitted levels.

Pause for thought

To what extent does the introduction of tradable permits lead to a lower level of total pollution (as opposed to its redistribution)?

How much can we rely on governments?

If governments are to be relied upon to set the optimum green taxes or regulations, several conditions must be met.

First, they must have the will to protect the environment. But governments are accountable to their electorates and must often appease various pressure groups, such as representatives of big business. In the USA, for example, there has been great resistance to cuts in greenhouse gases from the automobile, power and various other industries, many of which have powerful representation in Congress. One of the problems is that many of the environmental effects of our actions today will be on future generations, but governments represent today's generation, and today's generation may not be prepared to make the necessary sacrifices. This brings us back to the importance of education.

Secondly, it must be possible to identify just what the optimum is. This requires a knowledge of just what the environmental effects are of various activities, such as the emission of CO_2 into the atmosphere, and putting a monetary value on the economic, social and environmental impacts.

Finally, there is the problem that many environmental issues are global and not just local or national. KI 21 p 140 Many require concerted action by governments around the world. The history of international agreements on environmental issues, however, is plagued with difficulties between countries, which seem concerned mainly with their own national interests. Despite international agreements on restricting global warming, such as the Paris Agreement (2015/16) and Glasgow Climate Pact (2021), governments have been slow to adopt the necessary measures to achieve the agreed climate targets.

Recap

1. The effects of population pressures and market failures have led to growing environmental degradation.
2. The market fails to achieve a socially efficient use of the environment because large parts of the environment are a common resource, because production or consumption often generates environmental externalities, because of ignorance of the environmental effects of our actions, and because of a lack of concern for future generations.
3. One approach to protecting the environment is to impose charges for using the environment or taxes per unit of output. The problem with these methods is in identifying the appropriate charges or tax rates, since these will vary according to the environmental impact.
4. Another approach is to use laws and regulations, such as making certain practices illegal or putting limits on discharges. This is a less sophisticated alternative to taxes or charges, but it is safer when the environmental costs of certain actions are unknown.
5. Education can help to change attitudes towards the environment and the behaviour of consumers and firms.
6. Tradable permits are a mix of regulations and market-based systems. Firms are allocated permits to emit a certain level of pollution and these can then be traded. A firm that can reduce its pollution relatively cheaply below its permitted level can sell this credit to another firm that finds it costlier to do so. The system is an efficient and administratively cheap way of limiting pollution to a designated level. It can, however, lead to pollution being concentrated in certain areas and can reduce the pressure on firms to find cleaner methods of production.
7. Although governments can make a major contribution to reducing pollution, government action is unlikely to lead to the optimum outcome (however defined). Governments may be more concerned with short-run political considerations and will not have perfect information.

QUESTIONS

1. The table below gives the costs and benefits of an imaginary firm operating under perfect competition whose activities create a certain amount of pollution. (It is assumed that the costs of this pollution to society can be accurately measured.)

Output (units)	Price per unit (*MSB*) (£)	Marginal (private) costs to the firm (*MC*) (£)	Marginal external (pollution) costs (*MEC*) (£)	Marginal social costs (*MSC* = *MC* + *MEC*) (£)
1	100	30	20	50
2	100	30	22	52
3	100	35	25	60
4	100	45	30	75
5	100	60	40	100
6	100	78	55	133
7	100	100	77	177
8	100	130	110	240

   a. What is the profit-maximising level of output for this firm?
   b. What is the socially efficient level of output?
   c. Why might the marginal pollution costs increase in the way illustrated in this example?

2. In Figure 8.1 (page 196) the *MSC* curve is drawn as being steeper than the *MPC* curve. Under what circumstances would they be parallel?
3. Give additional examples of each of the four types of externality to those given on pages 196–8.
4. Give some examples of public goods (other than those given on page 199). Does the provider of these goods or services (the government or local authority) charge for their use? If so, is the method of charging based on the amount of the good people use? Is it a good method of charging? Could you suggest a better method?
5. Distinguish between publicly provided goods, public goods and merit goods.
6. Name some goods or services provided by the government or local authorities that are not public goods.
7. Some roads could be regarded as a public good, but some could be provided by the market. Which types of road could be provided by the market? Why? Would it be a good idea?
8. Assume that you have decided to buy an Ultra HD television. How do you set about ensuring that you make the right choice between the available makes?
9. Assume that you wanted the information given in (a)–(h) below. In which cases could you (i) buy perfect information; (ii) buy imperfect information; (iii) obtain information without paying for it; (iv) not obtain information?
   a. Which washing machine is the more reliable?
   b. Which of two jobs that are vacant is the more satisfying?
   c. Which builder will repair my roof most cheaply?
   d. Which builder will make the best job of repairing my roof?
   e. Which builder is best value for money?
   f. How big a mortgage would it be wise for me to take out?
   g. What course of higher education should I follow?
   h. What brand of washing powder washes whiter?

   In which cases are there non-monetary costs to you of finding out the information? How can you know whether the information you acquire is accurate or not?
10. The food industry provides a great deal of information about its products. Why, despite this, does the government run various campaigns about healthy eating?
11. Assume that a country had no state education at all. For what reasons might the private education system not provide the optimal allocation of resources to and within education?
12. What are the possible arguments in favour of fixing prices (a) below and (b) above the equilibrium? Are there any means of achieving the same social goals without fixing prices?
13. Assume that a firm discharges waste into a river. As a result, the marginal social costs (*MSC*) are greater than the firm's marginal (private) costs (*MC*). The table below shows how *MC, MSC, AR* and *MR* vary with output. Assume that the marginal private benefit (*MB*) is given by the price (*AR*). Assume also that there are no externalities on the consumption side, and that therefore *MSB* = *MB*.

Output	1	2	3	4	5	6	7	8
MC	23	21	23	25	27	30	35	42
MSC	35	34	38	42	46	52	60	72
TR	60	102	138	168	195	219	238	252
AR	60	51	46	42	39	36.5	34	31.5
MR	60	42	36	30	27	24	19	14

   a. How much will the firm produce if it seeks to maximise profits?
   b. What is the socially efficient level of output (assuming no externalities on the demand side)?
   c. How much is the marginal external cost at this level of output?
   d. What size of tax would be necessary for the firm to reduce its output to the socially efficient level?
   e. Why is the tax less than the marginal externality?
   f. Why might it be equitable to impose a lump-sum tax on this firm?
   g. Why will a lump-sum tax not affect the firm's output (assuming that in the long run the firm can still make at least normal profit)?
14. On a diagram similar to Figure 8.7 (on page 207), demonstrate how a subsidy can correct for an external benefit.
15. Why might it be better to ban certain activities that cause environmental damage rather than to tax them?
16. To what extent could property rights (either public or private) be successfully extended and invoked to curb the problem of industrial pollution (a) of the atmosphere, (b) of rivers, (c) by the dumping of toxic waste, (d) by the

erection of ugly buildings and (e) by the creation of high levels of noise?

17. What protection do private property rights in the real world give to sufferers of noise (a) from neighbours, (b) from traffic, (c) from transistor radios at the seaside?
18. How suitable are legal restrictions in the following cases?
    a. Ensuring adequate vehicle safety (e.g. that tyres have sufficient tread or that the vehicle is roadworthy).
    b. Reducing traffic congestion.
    c. Preventing the abuse of monopoly power.
    d. Ensuring that mergers are in the public interest.
    e. Ensuring that firms charge a price equal to marginal cost.
19. How would you evaluate the following?
    a. The external effects of building a reservoir in an area of outstanding natural beauty.
    b. The external effects of acid rain pollution from a power station.
20. Many economists have argued that a form of 'congestion tax' ought to be imposed on motorists who use their cars on busy roads, to take account of the external costs they impose on other road users and pedestrians. Compare the relative advantages and disadvantages of the following measures:
    a. Increasing the rate of duty on petrol.
    b. Increasing the annual road fund licence.
    c. Using a system such as that in Singapore (see Box 8.5), where charges are deducted from a pre-paid smart card inserted into a device in the car. Charges vary according to the time of day and/or the level of congestion.
    d. Installing cameras that record number plates of cars in a designated zone, and then fining their owners if a daily fixed fee for driving in the zone has not been paid (this system is used in London).
    e. Setting up toll booths to charge motorists for using certain stretches of road.
    f. The use of bus and cycle lanes at peak times.
    g. Subsidising public transport.
21. Give examples of how the government intervenes to protect the interests of dependants from bad economic decisions taken on their behalf.
22. What, do you think, has been the impact on car usage of free off-peak bus travel for senior citizens in England? Is there a case for extending the policy to (a) bus travel for senior citizens at all times of the day, as is the case in Scotland, Wales and Northern Ireland; (b) bus travel for all; (c) bus and train travel for senior citizens, as is the case in Northern Ireland and parts of Wales?
23. Make out a case for (a) increasing and (b) decreasing the role of the government in the allocation of resources.

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

8.1 **Corporate social responsibility.** An examination of social responsibility as a goal of firms and its effect on business performance.

8.2 **The police as a public service.** The extent to which policing can be classified as a public good.

8.3 **Deadweight loss from taxes on goods and services.** This shows the welfare loss from the imposition of a tax, which must be weighed against the redistributive and other gains from the tax.

8.4 **Regulating privatised industries.** This looks at the system of regulation applied to the major privatised industries in the UK.

8.5 **Libertarianism.** The views of the 'neo-Austrian' right that market capitalism has dynamic advantages in creating incentives to innovate and take risks.

8.6 **Public choice theory.** This examines how economists have attempted to extend their analysis of markets to the field of political decision-making.

8.7 **Perverse subsidies.** An examination of the use of subsidies around the world that are harmful to the environment.

8.8 **Selling the environment.** An examination of the Kyoto Protocol and successive attempts to reach international agreement on tackling climate change.

8.9 **A deeper shade of green.** Different approaches to environmental sustainability.

8.10 **Restricting car access to Athens.** A case study that examines how the Greeks have attempted to reduce local atmospheric pollution from road traffic.

8.11 **The right track for reform?** How successful has rail privatisation been in the UK?

8.12 **Can the market provide adequate protection for the environment?** This explains why markets generally fail to take into account environmental externalities.

8.13 **A Stern warning.** A summary of the most influential study of climate change in recent times: the Stern Review. This was an independent review led by Sir Nicholas Stern.

8.14 **Environmental auditing.** Are businesses becoming greener? A growing number of firms are subjecting themselves to an 'environmental audit' to judge just how 'green' they are.

8.15 **Evaluating new road schemes.** The system used in the UK of assessing the costs and benefits of proposed new roads.

8.16 **HS2: Is it really worth it?** The case for and against High Speed Rail in the UK.

8.17 **What price a human life?** Many projects involve saving lives, but how is a life to be evaluated in cost-benefit analysis?

WEB APPENDICES

8.1 **Private and social efficiency.** This appendix uses general equilibrium analysis to show how a perfect market economy in the absence of externalities will lead to a socially optimal allocation of resources.

8.2 **Cost–benefit analysis.** A technique to help decide whether or not a project should go ahead.

WEBSITES RELEVANT TO PART B

Numbers and sections refer to websites listed in the Websites Appendix at the end of the text and hotlinked from this book's website at www.pearsoned.co.uk/sloman.

- For news articles relevant to Part B, search for the *Sloman Economics News* site or follow the News Items link from the student website or MyLab Economics.
- For economic data, see sites in section B and particularly sites B1–6, 14, 21, 27, 29, 31-33, 39, 41, 48, 49. For data on the housing market, see sites B7–11.
- For student resources relevant to Part B, see sites C1–7, 9, 10, 14, 19–21, 25; D3, 19.
- For general news on markets, market failure and government intervention, see sites in section A, and particularly A1–5, 7–9, 18, 19, 21–26, 31, 35, 36. See also links to newspapers worldwide in A38, 39, 42–44; and see A41 for links to economics news articles from newspapers worldwide.
- For sites favouring the free market, see C17; E34. See also C18 for the development of ideas on the market and government intervention.
- For information on agriculture and the Common Agricultural Policy, see sites E14 and G9.
- For sites that look at companies, their scale of operation and market share, see D2; E4, 10; G7, 8.
- For information on poverty and inequality, see sites B18; E2, 9, 13, 32; G5.
- For information on taxes, benefits and the redistribution of income, see E9, 25, 30, 36; G5, 13.
- UK and EU departments relevant to competition policy can be found at sites E10; G7, 8.
- UK regulatory bodies can be found at sites E4, 11, 15, 16, 19, 22.
- For information on taxes and subsidies, see E25, 30, 36; G13. For use of green taxes, see E2, 14, 30; G11; H5.
- For information on health and the economics of health care, see E8; H8, 9.
- For policy on the environment and transport, see E2, 7, 11, 14, 29, 39; G10, 11, 19. See also H11.

Part C

Macroeconomics

Chapter 9

Aggregate demand and the business cycle

We turn now to *macroeconomics*. This will be the subject for this third part of the text and most of the final part.

In particular, we will be examining five key topics. The first is national output. What determines the size of national output? What causes it to grow? Why do growth rates fluctuate? Why do economies sometimes surge ahead and at other times languish in recession?

The second is employment and unemployment. What causes unemployment? If people who are unemployed want jobs, and if consumers want more goods and services, then why does our economy fail to provide a job for everyone who wants one?

The third is the issue of inflation. Why is it that the general level of prices always seems to rise, and only rarely fall? Why is inflation a problem? But why, if prices do fall, might that be a bad thing too? Why do countries' central banks, such as the Bank of England, set targets for the rate of inflation? And why is that target positive (e.g. 2 per cent) rather than zero?

The fourth issue is the financial system. We look at the role that financial institutions play in modern economies. In doing so, we analyse the financial crisis of the late 2000s, the initial responses of policy makers to limit the adverse impact on economies and the subsequent responses to try to prevent a similar crisis reoccurring.

The final topic, which is the subject of the final part of the text, concerns a country's economic relationships with other countries. We look at international trade and investment and at the flows of foreign currencies around the world.

In this chapter, after a preliminary look at the range of macroeconomic issues in Section 9.1, we focus in Section 9.2 on the first of these issues: national output. We consider why economies are inherently unstable and so experience the ups and downs of the 'business cycle'. We analyse the key purchasers of goods and services in the economy and the potential determinants of their spending.

We then use the circular flow model (Section 9.3) and simple Keynesian model (Section 9.4) to explain why spending may fluctuate and how these fluctuations can have a multiplied effect on the size of national income. We finish the chapter (Section 9.5) with a wide-ranging discussion of economic volatility, including an introduction to the role that the financial system plays in affecting short-term economic growth.

After studying this chapter, you should be able to answer the following questions:

- What are the key macroeconomic issues faced by all countries?
- What is meant by the business cycle?
- Who are the key groups of purchasers whose demands determine the total level of spending on a country's goods and services?
- What are the various flows of incomes around the economy? What causes these flows to expand or contract?
- How do we measure national output?
- What factors influence the level of spending by different sectors of the economy?
- What determines the level of national output at any one time?
- What is the effect on national income of an increase in spending?
- How do financial institutions and the availability of finance affect the business cycle?

9.1 INTRODUCTION TO MACROECONOMICS

What issues does macroeconomics tackle?

The first half of the text was concerned with microeconomics. We saw how it focuses on individual parts of the economy and with the demand and supply of particular goods and services and resources.

The issues addressed by macroeconomists, by contrast, relate in one way or another to the *total* level of spending in the economy (aggregate demand) or the *total* level of output (aggregate supply). Many of these issues are ones on which elections are won or lost. Is the economy growing and, if so, how rapidly? How can we avoid, or get out of, recessions? What causes unemployment and how can the rate be got down? Why is inflation a problem and what can be done to keep it at a modest level? What will happen to interest rates and if they were to change what would be their economic impact? How big a problem is government debt? Are banks lending too much or too little?

The questions we have just identified give you a flavour of the macroeconomic issues that we will be studying in the following chapters. We can group them under the following headings:

- economic growth
- unemployment
- inflation
- the economic relationships with the rest of the world
- the financial well-being of individuals, businesses and other organisations, governments and nations, and the relationship between the financial system and the economy.

We will be studying other issues too, such as consumer behaviour, finance and taxation, but these still link to these major macroeconomic issues and, more generally, to how economies function.

Key macroeconomic issues

Economic growth and the business cycle

The **rate of economic growth** measures the percentage change in an economy's output over a given period. If we measure it over a 12-month period, we are measuring the economy's annual rate of growth; if we measure it over a 3-month period we are measuring the quarterly rate of growth.

To be able to measure how quickly an economy is growing we need a means of measuring the value of a nation's output. The measure we use is *gross domestic product (GDP)*. (The Appendix to this chapter explains how it is calculated.) However, as we saw in Chapter 1 (page 17), to be able to compare changes in output from one period to the next we must eliminate those changes in GDP which result simply from changes in prices. In other words, we use *real* rather than *nominal* GDP figures to analyse changes in the volume of output.

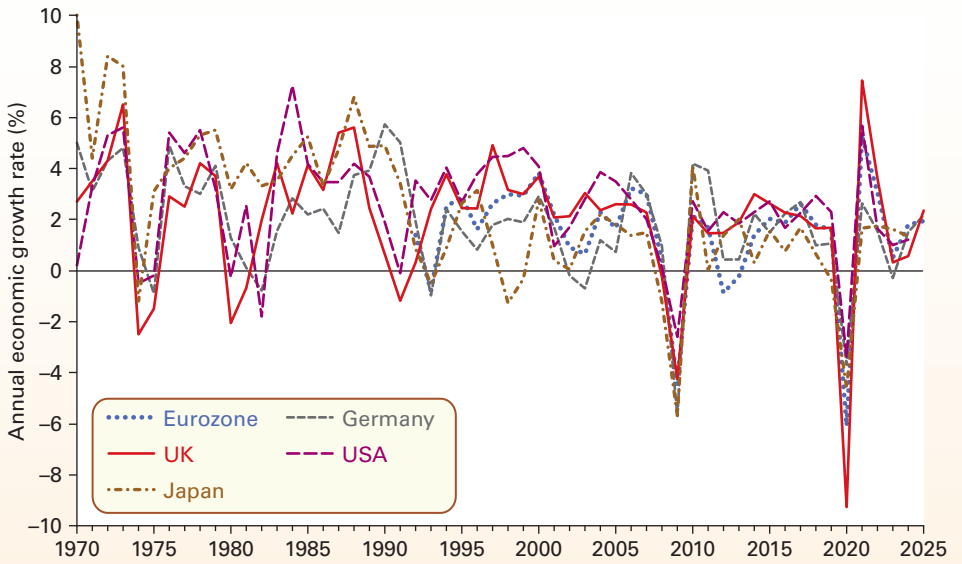
TC 5 p 17

Figure 9.1 shows the annual rates of economic growth – the annual percentage changes in real GDP – for a sample of economies. As you can see, countries rarely experience stable growth. Thus, while countries have experienced extraordinary volatility in recent times with the global financial crisis and then the COVID-19 pandemic, instability is an inherent feature of the macroeconomic environment. This is why, as we saw in Chapter 1, the instability of economies is a threshold concept. The resulting cycles in levels of economic activity are known as the **business cycle (or trade cycle)**. In some periods the economy will booming; in others economic growth will be low or even negative.

TC 3 p 10

Figure 9.1 illustrates the volatility of growth in both national economies, such as the UK and the USA, and also the eurozone. The business cycle extends beyond

Figure 9.1 Annual economic growth rates



Notes: Data from 2022 based on forecasts; eurozone figures are the weighted average of the 19 countries (as of 2022) using the euro; figures for Germany based on West Germany alone up to 1991

Source: Based on data in *AMECO Database* (European Commission) up to 1980; and *IMF World Economic Outlook Database*, IMF (October 2022)

country borders to the global economy, giving an **international business cycle**. Thus countries' business cycles have both a national and a global dimension. As the global dimension has increased in importance there has been a growing synchronicity of countries' business cycles.

Longer-term economic growth

Although growth rates fluctuate, most economies experience positive growth over the longer term. In other words, most economies have output paths that trend upwards over time. This can be seen in Figure 9.2 which plots the *levels* of real GDP and hence the output paths over time for the same economies analysed in Figure 9.1.

Figure 9.2 also highlights differences in the longer-term rates of growth of economies. For growth to be sustained over the longer term, an economy's capacity must increase. Hence, differences in long-term economic growth rates reflect differences in the growth of the productive capacity of economies. We explore the possible reasons behind these differences in Chapter 10.

Definitions

Rate of economic growth The percentage increase in output between two moments of time, typically over a 12-month period.

Business cycle or trade cycle The period fluctuations of a country's economic growth rates around the long-term trend.

International business cycle The tendency for groups of economies and the global economy to experience synchronised fluctuations in economic growth rates.

Unemployment

The inherent instability of economies has implications for the number of people in work and so for the number unable to find work. After all, higher levels of economic activity will tend to decrease unemployment numbers, while reduced economic activity will tend to increase them.

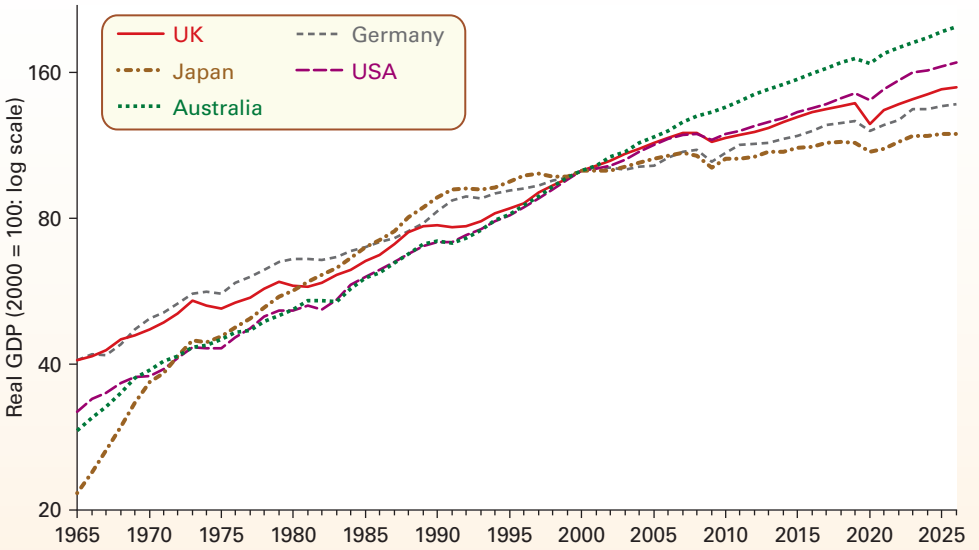
Unemployment numbers reflect more than just the position in the business cycle. For example, many countries have seen significant effects on their labour markets from rapid industrial change, technological advance and globalisation. While some new jobs are created, others are lost. Many people made redundant find they are not qualified for the new jobs being created.

Maximising employment opportunities and reducing unemployment is a key macroeconomic aim of governments, not only for the sake of the unemployed themselves, but also because unemployment represents a waste of human resources and because unemployment benefits are a drain on government revenues.

Measuring unemployment. Unemployment can be expressed either as a number (e.g. $^{1}/_{12}$ million) or as a percentage (e.g. 5 per cent).

The most usual definition that economists use for the **number unemployed** is: *those of working age who are without work, but who are available for work at current wage rates*.

Figure 9.2 Output paths, 1980 to 2020



Notes: Data from 2022 based on forecasts; eurozone figures are the weighted average of the 19 countries (as of 2022) using the euro; figures for Germany based on West Germany alone up to 1991

Source: Based on data in *AMECO Database* (European Commission) up to 1980; *IMF World Economic Outlook Database,* IMF (October 2022)

If the figure is to be expressed as a percentage, then it is a percentage of the total **labour force**. The labour force is defined as *those in employment plus those unemployed*. Thus if 30 million people were employed and 1.5 million people were unemployed, the **unemployment rate** would be:

$$\frac{1.5}{30 + 1.5} \times 100 = 4.76\%$$

Historical experience of unemployment. Unemployment rates in the 1980s and early 1990s were significantly higher in many advanced economies than in the previous three decades. Then, in the late 1990s and early 2000s, rates fell in some countries, such as the UK and USA, as captured in Figure 9.3.

Then, with the global economic crisis of the late 2000s, many countries experienced rising rates of unemployment. They then eased, before rising in response to the COVID-19 pandemic. However, many governments introduced policies to support employment, such as furloughing staff. This kept unemployment rates below levels seen during the global financial crisis.

In the UK, in recent years, there has been a move towards more flexible contracts, with many people's wages not keeping up with inflation and many working fewer hours than they would like. This has helped to reduce the rate of unemployment, but has created a problem of **underemployment**.

Inflation

By inflation we mean a general rise in prices throughout the economy. Government policy here is to keep inflation both low and stable. One of the most important reasons for this is that it will aid the process of economic decision making. For example, businesses will be able to set prices and wage rates, and make investment decisions with far more confidence.

Generally inflation tends to rise in periods of rapid economic growth: firms respond to the higher demand, partly by raising output, but partly also by raising prices. Conversely, in a recession, inflation is likely to fall: firms, faced with falling demand and rising stocks are likely to be unwilling to raise prices and may even cut them.

However, inflationary pressures can occur at any point in the business cycle from problems of supply. Examples include rising energy prices and transportation costs, and the shortage of key inputs. This was the situation that occurred with the supply-chain problems as economies began to 'open up' from the pandemic in 2021–22 and as oil and gas prices rose with

Definitions

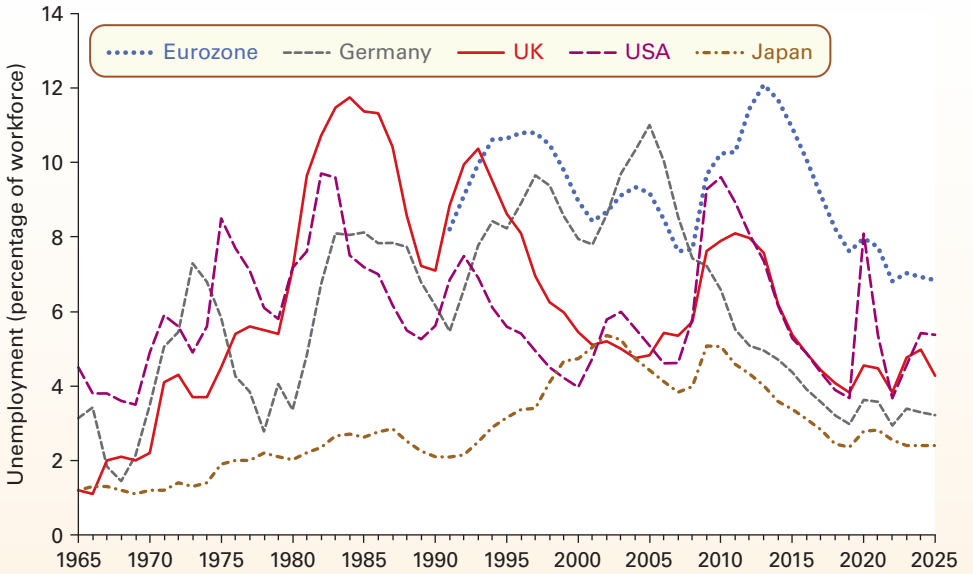
Number unemployed (economist's definition) Those of working age who are without work, but who are available for work at current wage rates.

Labour force The number employed plus the number unemployed.

Unemployment rate The number unemployed expressed as a percentage of the labour force.

Underemployment When people work fewer hours than they would like at their current wage rate.

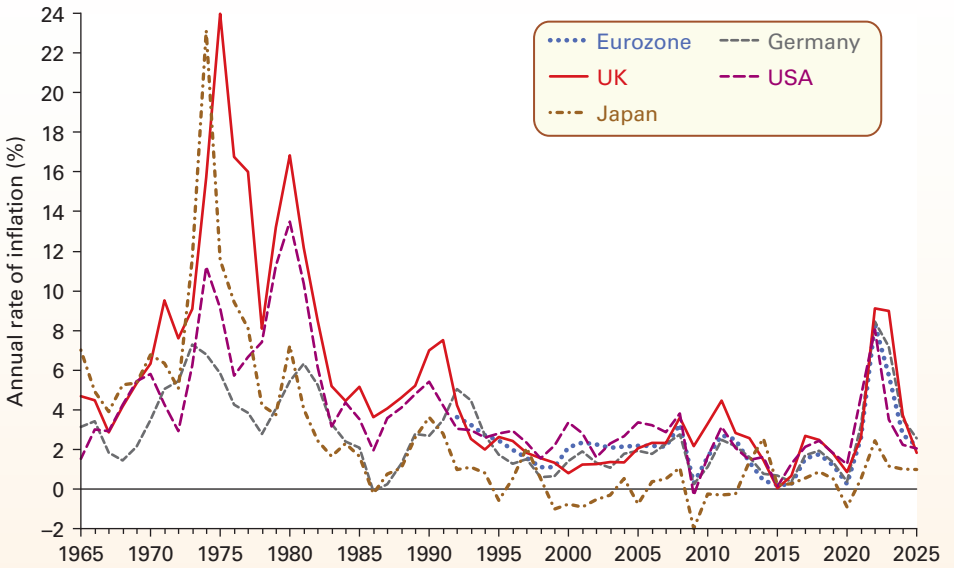
Figure 9.3 Standardised unemployment rates



Notes: Data from 2022 based on forecasts; eurozone figures are the weighted average of the 19 countries (as of 2022) using the euro; figures for Germany based on West Germany alone up to 1991

Source: Based on data in *AMECO Database* (European Commission) up to 1980; and *World Economic Outlook* (IMF, October 2022)

Figure 9.4 Inflation rates



Notes: Inflation rate is the annual percentage change in consumer prices; data from 2022 based on forecasts; eurozone figures are the weighted average of the 19 countries (as of 2022) using the euro; figures for Germany based on West Germany alone up to 1991

Source: Based on data in *AMECO Database* (European Commission) up to 1980; and *World Economic Outlook* (IMF, October 2022)

the invasion of Ukraine by Russia. Thus, while most countries had become used to low **inflation rates**, they now began to rise sharply, with a 'cost of living crisis' caused by wage growth failing to keep pace with prices. Nonetheless, rates of consumer price inflation remained much lower than in the past; in 1975, UK inflation reached 24 per cent (see Figure 9.4).

In most advanced economies today, **central banks**, such as the Bank of England, have a target for the rate of inflation that they are charged with meeting. In the UK the target for the growth of consumer prices is 2 per cent. The Bank of England then adjusts interest rates to try to keep inflation on target (we see how this works in Chapter 13).

Foreign trade and global economic relationships

A county's macroeconomic environment is shaped not only by domestic conditions but also by its economic relationships with other countries. These relationships evolve as the global economy develops and the world order changes. Take, for example, the rapid economic growth in economies such as China and India. This has had a major effect on patterns of world trade and development.

International economic relationships also evolve as countries or groups of countries come together to shape their economic relationships with other economies. With the EU–UK Trade and Cooperation Agreement, which took effect on 31 December 2020, a new set of economic relationships between the UK, the EU 27 and other foreign partners was to develop.

The balance of payments. One way of viewing the economic relationship between a country and other economies is through its **balance of payments account**. This records all transactions between the residents of that country and the rest of the world. These transactions enter as either debit items or credit items. The debit items include all payments *to* other countries: these include the country's purchases of imports, the spending on investment it makes abroad and the interest and dividends paid to people abroad who have invested in the country. The credit items include all receipts *from* other countries: from the sales of exports, from inward investment expenditure and from interest and dividends earned from abroad.

The sale of exports and any other receipts from abroad earn foreign currency. The purchase of imports or any other payments abroad use up foreign currency. If we start to spend more foreign currency than we earn, one of two things must happen. Both are likely to be a problem.

- *The balance of payments will go into deficit*. In other words, there will be a shortfall of foreign currencies. The government will therefore have to borrow money from abroad, or draw on its foreign currency reserves to make up the shortfall. This is a problem because, if it goes on too long, overseas debts will mount, along with the interest that must be paid; and/or reserves will begin to run low.
- *The exchange rate will fall*. The **exchange rate** is the rate at which one currency exchanges for another. For example, the exchange rate of the pound into the dollar might be £1 = $1.35.

If the government does nothing to correct the balance of payments deficit, then the exchange rate must fall: for example, to $1.30 or $1.25, or lower. (We will show just why this is so in Chapter 15.) A falling exchange rate is a problem because it pushes up the price of imports and may fuel inflation. This was the experience of the UK in the aftermath of the vote to leave the European Union, when the pound fell sharply. Also, if the exchange rate fluctuates, this can cause great uncertainty for traders and can damage international trade and economic growth.

Financial well-being

The financial system is an integral part of most economies. Financial markets, financial institutions and financial products have become increasingly important in determining the economic well-being of nations, organisations, government and people. The increasing importance of the financial system in everyday lives and to economies is known as **financialisation**.

One indicator of financialisation is the extent to which many of us now interact with financial institutions and make use of financial products. Financialisation is perhaps most frequently associated with the levels of indebtedness of individuals, businesses and organisations to financial institutions.

The importance of financial stability and the problem of financial distress. It is important for policy makers to ensure the stability of the financial system and the general financial well-being of **economic agents** (people, firms, government, etc.) This importance was most starkly demonstrated by the events surrounding the financial crisis of 2007–9, when many banks looked as

Definitions

Inflation rate (annual) The percentage increase in prices over a 12-month period.

Central bank a country's central bank is banker to the government and the banks as a whole (see Section 11.2). In most countries the central bank operates monetary policy by setting interest rates and influencing the supply of money. The central bank in the UK is the Bank of England; in the eurozone it is the European Central Bank (ECB) and in the USA it is the Federal Reserve Bank (the 'Fed').

Definitions

Balance of payments account A record of the country's transactions with the rest of the world. It shows the country's payments to or deposits in other countries (debits) and its receipts or deposits from other countries (credits). It also shows the balance between these debits and credits under various headings.

Exchange rate The rate at which one national currency exchanges for another. The rate is expressed as the amount of one currency that is necessary to purchase *one unit* of another currency (e.g. €1.15 = £1 and $1.35 = £1).

if they might become bankrupt. The crisis illustrated how the financial distress of financial institutions can lead to global economic turmoil. Because of the global interconnectedness of financial institutions and markets, problems can spread globally like a contagion.

And it was not just financial institutions that were distressed in the late 2000s, we also witnessed financially distressed households and businesses, many of whom were burdened by unsustainable levels of debt.

Subsequently, financial distress was to affect government too, especially in advanced economies. Governments were burdened by growing levels of debt as they spent more to offset rapidly weakening private-sector spending. At the same time, tax revenues fell because of lower or even negative economic growth. The consequence was a prolonged period during which many governments felt it necessary to tighten their budgets. And this constraint on government spending was to put a further brake on economic growth.

Financial accounts. In analysing financial well-being or distress, there are three key accounts which can be considered. These are compiled for the main sectors of the economy: the household, corporate and government sectors, and the whole economy.

- First, there is the *income account* which records the various *flows* of income (a credit) alongside the amounts either spent or saved (debits). Economic growth refers to the annual real growth in a country's income flows (i.e. after taking inflation into account).
- Next, there is the *financial account*. There are two elements here. First, we can record financial *flows*, which determine the net acquisition of financial wealth by each sector. These comprise new saving, borrowing or repayments. Reductions in the flows of borrowing, in countries like the UK and USA, were very important in explaining the credit crunch and subsequent deep recession of the late 2000s/early 2010s.

KI 23 p 173

The other element of the financial account is its **balance sheet**. A balance sheet is a record of *stocks* of **assets** and **liabilities** of individuals or institutions. An asset is something owned by or owed to you. A liability is a debt: i.e. something you owe to someone else. In the case of the financial account, we have a complete record of the stocks of financial assets (arising from saving) and financial liabilities (arising from borrowing) of a sector, and include things such as currency, bank deposits, loans, bonds and shares. The flows of borrowing during the 2000s meant that many individuals and organisations experienced a significant increase in stocks of financial liabilities.

KI 23 p 173

- Finally, there is the *capital account*, which records flows and stocks of *physical* assets and liabilities. Again, there are two elements. The first records the capital *flows* of the various sectors, which occur when acquiring or disposing of physical assets, such as property and machinery. The second records the *stock* of physical wealth held by the various sectors.

The national balance sheet. This is a measure of the wealth of a country (i.e. the nation's financial and physical stock of net assets). It shows the *composition* of a country's wealth and the contribution of each of the main *sectors* of the economy.

The balance of a sector's or country's stock of financial and non-financial assets over its financial liabilities is referred to as its **net worth**. An *increase* in the net worth of the sectors or the whole country implies greater financial well-being. However, during the 2000s many sectors experienced increases in net worth as asset values rose, despite the rising stock of financial liabilities. Subsequently, the increase in the stock of liabilities was not financially sustainable and asset prices fell.

Figure 9.5 is based on the national balance sheet for the UK since 1995. It shows the country's stock of net worth, including its value relative to annual national income (GDP). In 2020, the net worth of the UK was £10.7 trillion, equivalent to 5 times the country's annual GDP. The stock of net worth fell in both 2008 and 2009 at the height of the financial crisis and the economic slowdown. This reduced the country's net worth by nearly 9 per cent. It then fell again shortly afterwards in 2012.

These various accounts are part of an interconnected story detailing the financial well-being of a country's households, companies and government. To illustrate how, consider what would happen if, over a period of time, you were to spend more than the income you receive. This would result in your income account deteriorating.

To finance your excess spending, you could perhaps draw on savings or borrow from a bank. Either way, your financial balance sheet will deteriorate. Or you may dispose of some physical assets, such as property,

Definitions

Financialisation A term used to describe the process by which financial markets, institutions and instruments become increasingly significant in economies.

Economic agents The general term for individuals, firms, government and organisations when taking part in economic activities, such as buying, selling, saving, investing or in any other way interacting with other economic agents.

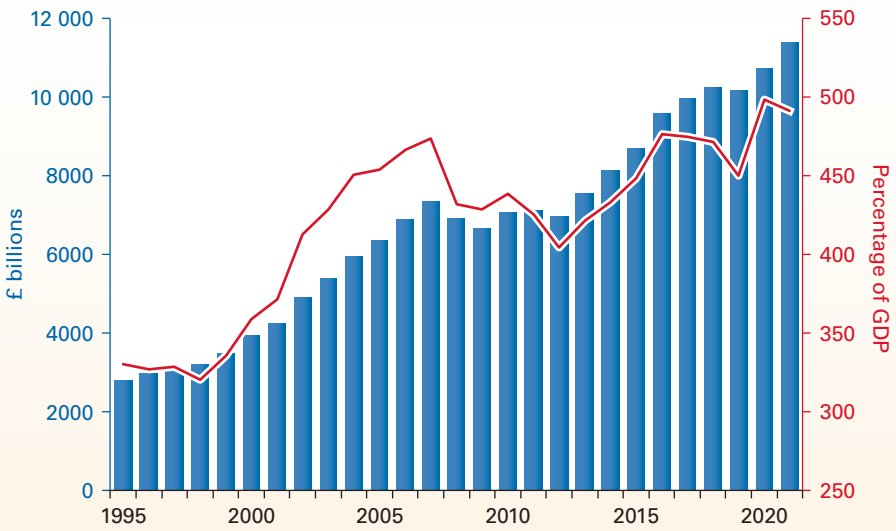
Balance sheet A record of the stock of assets and liabilities of an individual or institution.

Asset Possessions of an individual or institution, or claims held on others.

Liability Claims by others on an individual or institution; debts of that individual or institution.

Net worth The market value of a sector's stock of financial and non-financial wealth.

Figure 9.5 Net worth of the UK



Source: Based on series *CGDA* and *YBHA* (ONS)

KEY IDEA 32

Balance sheets affect people's behaviour. **The size and structure of governments', institutions' and individuals' liabilities (and assets too) affect economic well-being and can have significant effects on behaviour and economic activity.**

Pause for thought

Is the balance of payments account an income and expenditure account or a balance sheet?

in which case your capital balance will deteriorate. However excess spending is financed, net worth declines.

The importance of balance-sheet effects in influencing behaviour and, hence, economic activity has been increasingly recognised by both economists and policy makers, especially since the financial crisis of 2007–9. KI 32 p 239 Understanding these effects and their consequences is crucial in devising the most appropriate policies.

Government macroeconomic policy

From the above issues we can identify a series of macroeconomic policy objectives that governments might typically pursue:

- High and stable economic growth.
- Low unemployment.
- Low rates of inflation.
- The avoidance of balance of payments deficits and excessive exchange rate fluctuations.
- A stable financial system and the avoidance of financially distressed sectors of the economy, including government.

Unfortunately, these policy objectives may conflict. KI 31 p 206 For example, a policy designed to accelerate the rate of economic growth may result in a higher rate of inflation, a balance of payments deficit and excessive lending. Governments are thus often faced with awkward policy choices, further demonstrating how societies face trade-offs between economic objectives (see Section 8.4).

Recap

1. Macroeconomics, like microeconomics, looks at issues such as output, employment and prices; but it looks at them in the context of the whole economy.
2. Economies are inherently volatile, as evidenced by fluctuations in short-term economic growth rates. These fluctuations cause an economy's output path to fluctuate, generating what economists call the business cycle.
3. Among the macroeconomic goals that are generally of most concern to governments are economic growth, reducing unemployment, keeping inflation low and stable, avoiding balance of payments and exchange rate problems, a stable financial system and the avoidance of excessively financially distressed economic agents (i.e. households, businesses and governments).
4. Unfortunately, these goals are likely to conflict. Governments may thus be faced with difficult policy choices.

9.2 ECONOMIC VOLATILITY AND THE BUSINESS CYCLE

Why are levels of economic activity so volatile?

TC 3 p 10

Economies are volatile, as evidenced by the volatility of growth. They experience not only periods of expansion but also periods when growth is negative – when output levels contract. For many, the defining feature of the business cycle is the absence of growth in times of *recession*. A recession is officially defined as a period of two or more consecutive quarters when the economy experiences declining volumes of output.

Actual and potential growth and the output gap

The published statistics on growth show actual growth: the percentage change in national output over a period of time. As we saw in Section 9.1, this is usually measured over a year (12 months) or a quarter (3 months).

We should be careful to distinguish between actual growth and potential growth. **Potential growth** is the speed at which the economy *could* grow. It is the percentage annual increase in the economy's *capacity* to produce: the rate of growth in *potential output*.

Potential output (i.e. potential GDP) is the level of output when the economy is operating at 'normal capacity utilisation'. This allows for firms having a planned degree of spare capacity to meet unexpected demand or for hold-ups in supply. It also allows for some unemployment as people move from job to job. Because potential output is normal-capacity output, it is somewhat below full-capacity output, which is the absolute maximum that could be produced with firms working flat out.

The difference between actual and potential output is known as the **output gap**. Thus if actual output exceeds potential output, the output gap is positive: the economy is operating above normal capacity utilisation. If actual output is below potential output, the output gap is negative: the economy is operating below normal capacity utilisation. Box 9.1 looks at the output gap since 1970 for five major industrial economies.

Assume that the actual growth rate is less than the potential growth rate. This will lead to an increase in spare capacity and probably an increase in unemployment. In turn, the output gap will become less positive or perhaps more negative depending on the economy's starting point.

In contrast, if the actual growth rate were to exceed the potential growth rate, there would be a reduction in spare capacity and the output gap would become less negative or more positive. However, periods when actual growth exceeds potential growth can only be temporary. In the long run, the actual growth rate will be limited to the potential growth rate.

The hypothetical business cycle

TC 3 p 10

Actual growth tends to fluctuate. In some years, countries will experience high rates of economic growth: the country experiences a boom. In other years, economic growth is low or even negative: the country experiences a slowdown or recession. This cycle of expansion and slowdown causes fluctuations in the path of output.

Figure 9.6 illustrates a hypothetical business cycle. While it is a stylised representation of the business cycle, it is useful for illustrating four identifiable 'phases' of the cycle.

1. *The upturn*. In this phase, a contracting or stagnant economy begins to recover, and growth in actual output resumes, or begins to accelerate.
2. *The expansion*. During this phase, there is rapid economic growth: the economy is booming. A fuller use is made of resources, and the gap between actual and potential output narrows.
3. *The peaking out*. During this phase, growth slows down or even ceases.
4. *The slowdown, recession or slump*. During this phase, there is little or no growth or even a decline in output. Increasing slack develops in the economy. The economy is operating with a negative output gap.

A word of caution: do not confuse a high *level* of output with a high rate of *growth* in output. The level of output is highest in phase 3. The rate of growth in output is highest in phase 2 (i.e. where the curve is steepest).

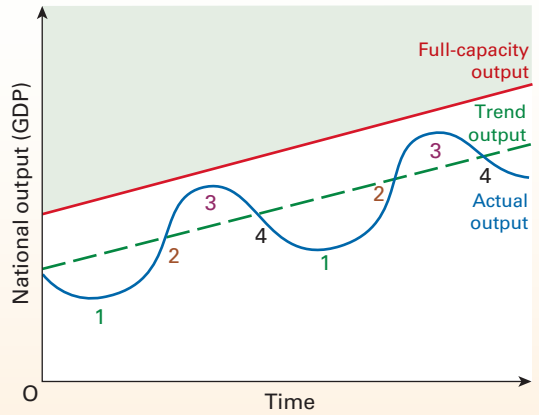
Definitions

Potential growth The percentage annual increase in the capacity of the economy to produce.

Potential output The economically sustainable level output that could be produced in the economy: i.e. one that involves a 'normal' level of capacity utilisation and does not result in rising inflation.

Output gap The difference between actual and potential output. When actual output exceeds potential output, the gap is positive. When actual output is less than potential output, the gap is negative.

Figure 9.6 A hypothetical business cycle



Pause for thought

Figure 9.6 shows a decline in actual output in recession. Redraw the diagram, only this time show a mere slowing down of growth in phase 4.

Long-term output trend. A line can be drawn showing the trend of national output over time (i.e. ignoring the cyclical fluctuations around the trend). This is shown as the dashed line in Figure 9.6. If, over time, firms on average operate with a 'normal' degree of capacity utilisation (a zero output gap), the trend output line will be the same as the potential output line. Also, if the average level of capacity that is unutilised stays constant from one cycle to another, the trend line will have the same slope as the full-capacity output line. In other words, the trend (or potential) rate of growth will be the same as the rate of growth of capacity.

If, however, the level of unutilised capacity changes from one cycle to another, then the trend line will have a different slope from the full-capacity output line. For example, if unemployment and unused industrial capacity *rise* from one peak to another, or from one trough to another, the trend line will move further away from the full-capacity output line (i.e. it will be less steep).

The business cycle in practice

The hypothetical business cycle illustrated in Figure 9.6 is nice and smooth and regular. Drawing it this way allows us to make a clear distinction between each of the four phases. In practice, however, business cycles are highly irregular. You can see this by examining Figure 9.1 (on page 234). They are irregular in two important ways.

The length of the phases. Some booms are short lived, lasting only a few months or so. Others are much longer, lasting several years. Likewise some recessions are short while others are long.

The magnitude of the phases. Sometimes in phase 2, there is a very high rate of economic growth, perhaps 4 per cent per annum or more. On other occasions in phase 2, growth is much gentler. Sometimes in phase 4 there is a recession, with an actual decline in output, as occurred in 2008–9. On other occasions, phase 4 is merely a 'pause', with growth simply being low.

Spending, output and the business cycle

The fluctuations in economic activity that characterise the business cycle reflect, in one way or another, fluctuations in total spending. In analysing the business cycle, it is therefore important to analyse the *sources* of volatility in spending and the *processes* or mechanisms by which this volatility works through the economy and affects output, inflation and employment. These processes may help to smooth the business cycle or, alternatively, amplify the peaks and troughs of the cycle.

Aggregate demand and its components

The focus of much of the analysis on the business cycle is on understanding fluctuations in **aggregate demand** (*AD*). This is the total spending on goods and services made within the country ('domestically produced goods and services'). Periods of rapid growth are associated with periods of rapid expansion of aggregate demand. Periods of recession are associated with a decline in aggregate demand.

Aggregate demand consists of spending by four groups of people: consumers on goods and services (*C*), firms on investment (*I*), the government on goods, services and investment (such as education, health and new roads) (*G*) and people abroad on this country's exports (*X*). From these four we have to subtract any imports (*M*) since aggregate demand refers only to spending on *domestic* production. Thus:[1]

$$AD = C + I + G + X - M$$

Table 9.1 illustrates aggregate demand and its components. It presents the average percentage composition of aggregate demand for a selection of countries over the period 1970 to 2020.

The first three columns show the volume of purchases made by each country's residents, whether on domestically produced goods or imports. To arrive at the figure for aggregate demand, we must add the consumption on each country's products by people abroad (exports), but also subtract that part of the expenditure in the first three columns going on goods and services from abroad (imports).

[1] An alternative way of specifying this is to focus on just the component of each that goes to domestic firms. We use a subscript 'd' to refer to this component (i.e. with the imported component subtracted). Thus $AD = C_d + I_d + G_d + X$.

BOX 9.1 OUTPUT GAPS AND THE BUSINESS CYCLE

An alternative measure of excess or deficient demand

If the economy grows, how fast and for how long can it grow before it runs into inflationary problems? On the other hand, what minimum rate must be achieved to avoid rising unemployment?

To answer these questions, economists have developed the concept of 'output gaps'.[1] As we have seen, the output gap is the difference between actual output and potential (normal-capacity) output.

If actual output is below potential output (the gap is negative), there will be a higher than normal level of unemployment as firms are operating below their normal level of capacity utilisation. There will, however, be a downward pressure on inflation, resulting from a lower than normal level of demand both for goods and services and for labour and other resources.

If actual output is above potential output (the gap is positive), there will be excess demand and a rise in the rate of inflation.

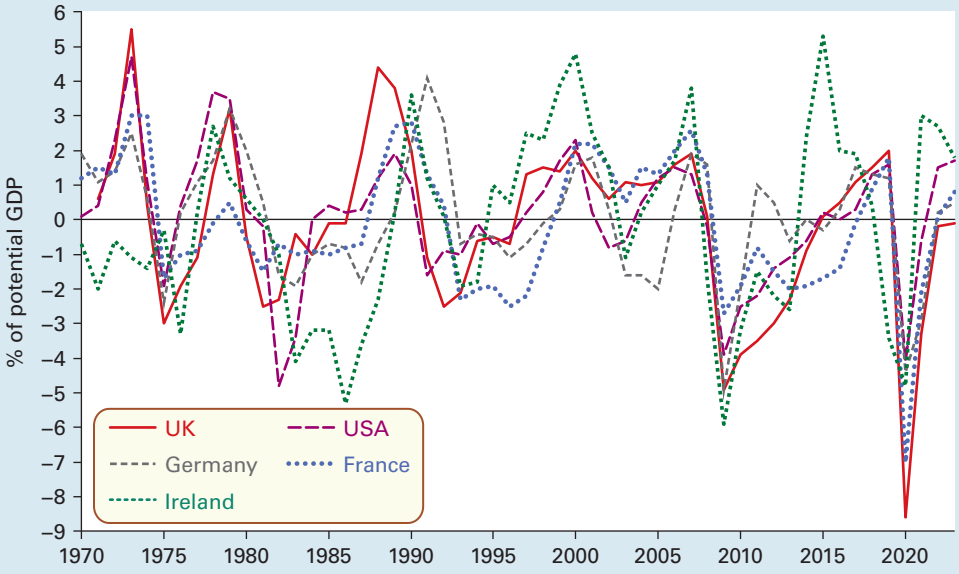
Generally, the gap will be negative in a recession and positive in a boom. In other words, output gaps follow the course of the business cycle.

But how do we measure output gaps? There are two principal statistical techniques.

De-trending techniques. This approach is a purely mechanical exercise which involves smoothing the actual GDP figures. In

[1]See Giorno et al., 'Potential output, output gaps and structural budget balances', *OECD Economic Studies*, no. 24, 1995: 1.

Output gaps



Note: Data from 2022 based on forecasts; figures for Germany based on West Germany alone up to 1991

Source: Based on data from *AMECO Database* (European Commission, DGECFIN, May 2022)

The balance of exports over imports is referred to as *net exports* or the **balance of trade in goods and services** or simply the **balance of trade.** A balance of trade *surplus* occurs when exports exceed imports while a *deficit* occurs when imports exceed exports. We look at the balance of trade in the context of the overall flows of money between residents of a country and the rest of the world – the balance of payments account – in Chapter 15.

Although the composition of aggregate demand varies across countries, spending by consumers on goods and services (*C*) is, in most countries, the largest component of aggregate demand.

Pause for thought

Why are the figures for exports and imports so high for Singapore and relatively high for Ireland and the Netherlands, but so low for the USA and Brazil?

Aggregate supply

While economic analysis of the business cycle typically stresses the importance of fluctuations in aggregate demand, fluctuations in **aggregate supply** can also cause volatility. Sudden sharp changes to input prices, such as in the price of oil, can be one such cause.

EXPLORING ECONOMICS

doing this, it attempts to fit a trend growth path along the lines of the dashed line in Figure 9.6. The main disadvantage of this approach is that it is not grounded in economic theory and therefore does not take into consideration those factors likely to determine normal-capacity output.

Production function approach. Many institutions, such as the European Union, use an approach which borrows ideas from economic theory. Specifically, this uses the idea of a production function which relates output to a set of inputs. Estimates of potential output are generated by using statistics on the size of a country's capital stock (see Box 10.3), the potential available labour input and, finally, the productivity or effectiveness of these inputs in producing output.

In addition to these statistical approaches use could be made of *business surveys*. In other words, we ask businesses directly. However, survey-based evidence can provide only a broad guide to rates of capacity utilisation and whether there is deficient or excess demand.

International evidence

The chart shows output gaps for six countries from 1970 estimated using a production function approach. What is apparent from the chart is that all the countries have experienced significant output gaps, both positive and negative. This is consistent with a theme that we shall see throughout the second half of the text: economies are inherently volatile. In other words, countries experience chart cycles.

TC 3 p 10

The chart does show that the characteristics of countries' business cycles can differ, particularly in terms of depth and duration. But, we also see evidence of an international business cycle (see pages 233–4) where national cycles appear to share characteristics. This global component of countries' business cycles is particularly stark in the late 2000s and early 2010s. Increasing global interconnectedness from financial and trading links meant that the financial crisis of the late 2000s spread like a contagion.

While output gaps vary from year to year, over the longer-term the average output gap tends towards zero. As we can see from the table below, this means that for our selection of countries from 1970 the actual rate of economic growth is roughly the same as the potential rate.

Average annual growth in actual and potential output, % (1970–2023)

	Average annual growth rates (%)	
	Actual output (real GDP)	Potential output
France	2.19	2.16
Germany	1.96	2.07
Ireland	5.12	5.07
Netherlands	2.29	2.30
UK	2.11	2.09
USA	2.72	2.73

Source: *AMECO Database* (European Commission, DGECFIN)

1. Characterise the state of an economy during positive and negative output gaps.
2. Are all business cycles the same?

Using the AMECO database, calculate the average output gap for France, Germany, Ireland, UK and USA for the period 2009–11 and then use their values to construct a column chart. Briefly summarise your findings.

Meanwhile, the COVID-19 pandemic saw many governments implement health intervention measures, such as lockdowns, that resulted in negative shocks to both aggregate demand *and* aggregate supply.

Some economists go further and argue that shifts in aggregate supply are the *primary* source of economic volatility. They argue that economies are frequently affected by supply 'shocks'. These could include technological shocks, or changes in the regulatory climate or the political environment, such as a war. These, in turn, affect the economy's potential output, either negatively or positively. Consequently, the business cycle is the result of fluctuations in potential output, which in turn affect actual output. A business cycle resulting from fluctuations in aggregate supply is known as a **real business cycle.**

Definitions

Aggregate demand Total spending on goods and services produced in the economy. It consists of four elements, consumer expenditure (C), investment (I), government expenditure (G) and the expenditure on exports (X), less any expenditure on foreign goods and services (M). Thus $AD = C + I + G + X - M$.

Balance on trade in goods or balance of visible trade or merchandise balance Exports of goods minus imports of goods.

Aggregate supply The total amount that firms plan to supply at any given level of prices.

Real business cycle Fluctuations in economic activity explained in terms of shifts in aggregate supply, rather than aggregate demand.

Table 9.1 Composition of aggregate demand, % (average 1970–2020)

	Household final consumption	Gross capital formation (public and private)	General government final consumption	Exports	Imports	Net exports (*X*–*M*)
Australia	56.5	26.6	17.8	17.8	18.4	−0.7
Brazil	64.8	19.8	15.8	10.4	10.4	0.0
China	45.1	38.9	14.2	16.9	15.2	1.7
France	54.7	23.2	22.1	24.1	24.1	0.0
Germany	56.0	23.5	19.6	29.0	27.5	1.5
India	66.7	27.1	10.5	12.2	14.1	−1.9
Ireland	52.7	24.7	17.5	72.7	66.6	6.1
Japan	53.1	29.9	16.1	12.7	11.8	0.9
Netherlands	49.1	22.3	22.7	60.5	54.4	6.1
Singapore	45.0	33.3	10.2	175.5	163.9	11.5
Sweden	47.8	24.9	25.2	35.7	32.8	2.9
UK	66.2	20.1	19.5	26.0	26.8	−0.9
USA	64.6	22.0	15.6	9.7	11.9	−2.2

Note: Based on constant-price data

Source: *National Accounts Main Aggregates Database* (United Nations Statistics Division)

Recap

1. Actual growth must be distinguished from potential growth. The actual growth rate is the percentage annual increase in the output that is actually produced, whereas potential growth is the percentage annual increase in the normal capacity of the economy to produce (whether or not it is actually produced).
2. Actual growth will fluctuate with the course of the business cycle. The cycle can be broken down into four phases: the upturn, the expansion, the peaking out and the slowdown or recession. In practice, the length and magnitude of these phases will vary: the cycle is thus irregular.
3. Explanations of the business cycle tend to focus on fluctuations in aggregate demand. This requires a deeper understanding of the behaviour of the components of aggregate demand, including that of private-sector behaviour.
4. However, fluctuations in aggregate supply can also result in economic instability. Some economists go further and argue that business cycles can result from fluctuations in potential output caused by frequent technology and other shocks.

9.3 THE CIRCULAR FLOW OF INCOME MODEL

What determines the size of flows of income to firms?

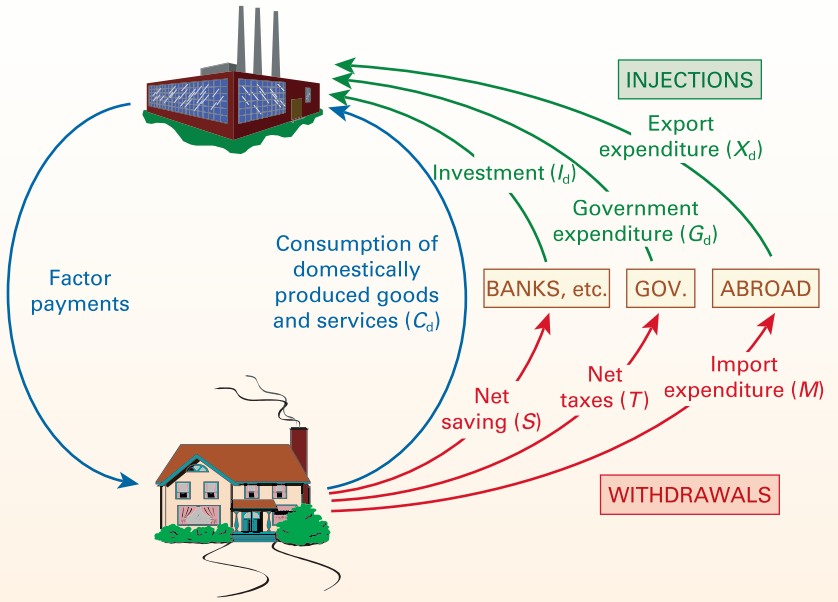
TC 4 p 13

A useful model for showing the relationship between aggregate demand and output is the *circular flow of income model*, which is illustrated in Figure 9.7. It is an extension of the model we looked at back in Chapter 1 (see Figure 1.5 on page 16).

From the left-hand side of the diagram we can identify two major groups: *firms* and *households*. Each group has two roles. Firms are producers of goods and services; they are also the employers of labour and other factors of production. Households (which include all individuals) are the consumers of goods and services; they are also the suppliers of labour and various other factors of production. In the diagram there is an inner flow and various outer flows of incomes between these two groups.

Before we look at the various parts of the diagram, a word of warning. Do not confuse *money* and *income*. Money is a stock concept. At any given time, there is a certain quantity of money in the economy (e.g. £1 trillion). But that does not tell us the level of

Figure 9.7 The circular flow of income



national *income*. Income is a flow concept, measured as so much *per period of time*.

The relationship between money and income depends on how rapidly the money *circulates*: its 'velocity of circulation'. (We will examine this concept in detail later on.) If there is £1 trillion of money in the economy and each £1 on average is paid out as income twice each year, then annual national income will be £2 trillion.

The inner flow, withdrawals and injections

The inner flow

Firms pay money to households in the form of wages and salaries, dividends on shares, interest and rent. These payments are in return for the services of the factors of production – labour, capital and land – that are supplied by households. Thus on the left-hand side of the diagram, money flows *directly* from firms to households as 'factor payments'.

Households, in turn, pay money to domestic firms with their **consumption of domestically produced goods and services** (C_d). This is shown on the right-hand side of the inner flow. There is thus a circular flow of payments from firms to households to firms and so on.

If households spend *all* their incomes on buying domestic goods and services, and if firms pay out *all* this income they receive as factor payments to domestic households, and if the velocity of circulation does not change, the flow will continue at the same level indefinitely. The money just goes round and round at the same speed and incomes remain unchanged.

Pause for thought

Would this argument still hold if prices rose?

In the real world, of course, it is not as simple as this – if it were, economies would not be characterised by the instability we are trying to understand. To help understand how incomes levels change we introduce two other groups of purchasers and suppliers into the model and also the financial system. While this increases the complexity of the model, it makes it more realistic and increases the ways in which the total spending on goods and services in the economy is affected.

The two additional groups are *government* and *overseas purchasers and suppliers*. The latter comprises the foreign equivalents of our three domestic groupings. For instance, it includes French households, Japanese car manufacturers and the American government.

The other element in the model is *financial institutions*, such as banks and building societies, and they

play a key role. These institutions provide the link between those who wish to borrow and those who wish to save. In other words, they are 'intermediaries', which allow some in the economy to save while others borrow. For instance, they can provide firms with the access to the credit they need to fund investment projects, such as the purchase of new machinery.

We now are ready to incorporate our three new elements: government, the overseas sector and the financial system into our model.

We begin by focusing on flows of income to these three groups from the inner flow. These are known as withdrawals from the inner flow.

Withdrawals (W)

Only part of the incomes received by households will be spent on the goods and services of domestic firms. The remainder will be withdrawn from the inner flow. Likewise, only part of the incomes generated by firms will be paid to UK households. The remainder of this will also be withdrawn. There are three forms of **withdrawals** (or 'leakages' as they are sometimes called).

Net saving (S). Saving is income that households choose not to spend but to put aside for the future. Savings are normally deposited in financial institutions such as banks and building societies. This is shown in the bottom centre of the diagram. Money flows from households to 'banks, etc.'. What we are seeking to measure here, however, is the net flow from households to financial institutions. We therefore have to subtract from saving any borrowing or drawing on past savings by households to arrive at the *net* saving flow. Of course, if household borrowing exceeded saving, the net flow would be in the other direction: it would be negative.

Net taxes (T). When people pay taxes (to either central or local government), this represents a withdrawal of money from the inner flow in much the same way as saving: only in this case, people have no choice!

Some taxes, such as income tax are paid out of household incomes. Others, such as Value Added Tax (VAT), are paid out of consumer expenditure. Other taxes, such as corporation tax, are paid out of firms' incomes before being received by households as dividends on shares. For simplicity, however, taxes are shown in Figure 9.7 as leaving the circular flow at just one point. It does not affect the argument.

When, however, people receive *benefits* from the government, such as unemployment benefits, child benefit and pensions, the money flows the other way. Benefits are thus equivalent to a 'negative tax'. These benefits are known as **transfer payments**. They transfer money from one group of people (taxpayers) to others (the recipients).

In the model, 'net taxes' (T) represent the *net* flow to the government from households and firms. It consists of total taxes minus benefits.

Import expenditure (M). Not all household consumption (C) is of totally home-produced goods (C_d). Households spend some of their incomes on imported goods and services, or on goods and services using imported components. Although the money that consumers spend on such goods initially flows to domestic retailers, it will eventually find its way abroad, either when the retailers or wholesalers themselves import them, or when domestic manufacturers purchase imported inputs to make their products. This expenditure on imports constitutes the third withdrawal from the inner flow. This money flows abroad.

As we shall see, households are not the only group to purchase imported goods and services or goods and services using imported components: firms and government do too. These expenditures also contribute towards the sum of import expenditures (M) and affect the level of aggregate demand.

Total withdrawals are simply the sum of net saving, net taxes and the expenditure on imports:

$$W = S + T + M$$

Injections (J)

Only part of the demand for firms' output arises from consumers' expenditure. The remainder comes from other sources outside the inner flow. These additional components of aggregate demand are known as **injections** ***(J)***. There are three types of injection.

Definition

The consumption of domestically produced goods and services (C_d) The direct flow of payments from households to firms for goods and services produced within the country.

Definitions

Withdrawals (W) (or leakages) Incomes of households or firms that are not passed on round the inner flow. Withdrawals equal net saving (S) plus net taxes (T) plus import expenditure (M): $W = S + T + M$.

Transfer payments Moneys transferred from one person or group to another (e.g. from the government to individuals) without production taking place.

Investment on domestically produced capital (I_d). This is firms' spending on domestically produced goods and services after obtaining the money from various financial institutions – either past savings or loans, or through a new issue of shares. They may invest in plant and equipment or may simply spend the money on building up stocks of inputs, semi-finished or finished goods.

Not all of the investment expenditure (I) undertaken by domestic firms is on totally home-produced items. Investment expenditure on capital equipment, stocks, etc. produced overseas contributes towards import expenditure (M).

Government purchases of domestically produced goods and services (G_d). When the government (both central and local) spends money on goods and services produced by domestic firms, this counts as an injection. Note that government expenditure in this model does *not* include state benefits and hence the use of the term 'government purchases'. Benefits, as we saw above, are transfer payments and are the equivalent of negative taxes and have the effect of reducing the T component of withdrawals.

As well as providing goods and services by purchasing from firms, governments own and run operations themselves. The government therefore becomes a purchaser on behalf of the public, for example of health and education services. The wages of public-sector staff will be a component of the government's expenditure and are a flow of factor payments to households.

As with investment, not all government purchases (G) are on totally home-produced goods and services. Expenditures on items made overseas contribute towards import expenditure (M).

Pause for thought

How might greater state involvement in the delivery of services like health and education affect the mix of C *and* G *in aggregate demand?*

Export expenditure (X_d). Money flows into the circular flow from abroad when households, firms and governments abroad buy our exports of goods and services. Note that, as with the other two injections, only those parts of exports made in the country should be counted. Any imported materials or components into the exports should be deducted.

Total injections are simply the sum of investment, government expenditure and exports, in each case with any imported component subtracted:

$$J = I_d + G_d + X_d$$

Aggregate demand, as we have seen, is the total spending on domestic firms. In other words, it is the spending by the household sector on domestically produced goods and services (C_d), plus the three injections:[2]

$$AD = C_d + J$$

The relationship between withdrawals and injections

There are indirect links between saving and investment, taxation and government expenditure, and imports and exports, via financial institutions, the government (central and local) and foreign countries respectively.

- If more money is saved, there will be more available for banks and other financial institutions to lend out.
- If tax receipts are higher, the government may be keener to increase its expenditure.
- Finally, if imports increase, the extra income generated abroad may stimulate demand for exports. Also, as we shall see in Chapter 15, the worsening balance of trade may push down the exchange rate, which will lower the foreign currency price of exports and thereby increase demand for them.

These links, however, do not guarantee that $S = I_d$, or $G_d = T$ or $M = X_d$.

Take investment and saving. The point here is that the decisions to save and invest are made by different people, and thus they plan to save and invest different amounts. Likewise, the demand for imports may not equal the demand for exports. As far as the government is concerned, it may choose not to spend all its tax revenues: to run a 'budget surplus'; or it may choose to spend more than it receives in taxes: to run a 'budget deficit' – by borrowing to make up the difference.

Thus planned injections (J) may not equal planned withdrawals (W).

Definition

Injections (J) Expenditure on the production of domestic firms coming from outside the inner flow of the circular flow of income. Injections equal investment (I_d) plus government expenditure (G_d) plus expenditure on exports (X).

[2]Note that this definition of aggregate demand ($AD = C_d + J$) is equivalent to the one we gave on page 241, i.e. $AD = C + I + G + X - M$, since both the terms C_d and J exclude expenditure on imports.

Equilibrium in the circular flow

The circular flow of income model helps us to understand how fluctuations in aggregate demand can cause fluctuations in national income. When injections do not equal withdrawals, a state of *disequilibrium* will exist: aggregate demand will rise or fall and so will national income. Disequilibrium results in a chain reaction that brings the economy back to a state of equilibrium where injections are equal to withdrawals.

Take the case where injections exceed withdrawals. Perhaps there has been a rise in business confidence and hence a rise in investment; or perhaps there has been a tax cut so that withdrawals have fallen and consumption rises. The level of expenditure will rise: there will be a rise in aggregate demand.

This extra spending will increase firms' sales and thus encourage them to produce more. Total output in the economy will rise. Moving anti-clockwise around the circular flow, this will result in firms paying out more in wages, salaries, profits, rent and interest (i.e. factor payments). In other words, national income will rise.

However, as national income rises, so households will not only spend more on domestic goods (C_d), but also save more (S), pay more taxes (T) and buy more imports (M). In other words, withdrawals will rise. This will continue until they have risen to equal injections. At that point, national income will stop rising, and so will withdrawals. Equilibrium has been reached where:

TC 8 p 37

$$W = J$$

But when aggregate demand changes *by how much* will national income change? What will the new equilibrium level of national income be? To help analyse this, the next section develops the 'simple Keynesian model' of national income determination. It provides us with a framework to examine the relationship between national income and the component parts of the circular flow of income: consumption, withdrawals and injections.

Recap

1. The circular flow of income model depicts the flows of money round the economy. The inner flow shows the direct flows between firms and households. Money flows from firms to households in the form of factor payments, and back again as consumer expenditure on domestically produced goods and services.
2. Not all incomes get passed on directly round the inner flow. Some is withdrawn in the form of net saving; some is paid in net taxes; and some goes abroad as expenditure on imports.
3. Likewise not all expenditure on domestic firms is by domestic consumers. Some is injected from outside the inner flow in the form of investment expenditure, government expenditure and expenditure on the country's exports.
4. The circular flow will be in equilibrium when planned injections equal planned withdrawals. But, planned injections and withdrawals are unlikely to be equal. This will result in a chain reaction that returns the economy to a position of equilibrium.

9.4 SIMPLE KEYNESIAN MODEL OF NATIONAL INCOME DETERMINATION

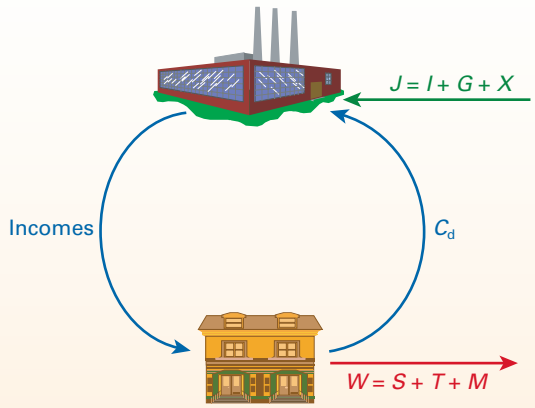
How do changes in aggregate demand affect national income?

The following analysis is based on the theory developed by John Maynard Keynes back in the 1930s, a theory that has had a profound influence on economics (see Case Studies 9.1 and 9.2 on the student website). Keynes argued that, without government intervention to steer the economy, countries could lurch from unsustainable growth to deep and prolonged recessions. In other words, government intervention has a role to play in reducing macroeconomic instability.

The central argument is that the level of production in the economy depends on the level of aggregate demand. If people buy more, firms will produce more in response to this, providing they have spare capacity. If people buy less, firms will cut down their production and lay off workers. But just *how much* will national income rise or fall as aggregate demand changes?

The simple Keynesian model can be explained most simply by returning to the circular flow of income diagram. Figure 9.8 shows a simplified version of the circular flow model that we looked at in the previous section.

Figure 9.8 Simplified version of the circular flow model



We saw in Section 9.3 that aggregate demand will be constant when the total level of injections (J) equals the total level of withdrawals (W). But, if injections do not equal withdrawals, a state of disequilibrium exists. What will bring them back into equilibrium is a change in national income and employment.

Start with a state of equilibrium, where injections equal withdrawals. Now assume that there is a *fall* in injections, perhaps because of a fall in government purchases. As a result, aggregate demand ($C_d + J$) will be lower. Firms will respond to the lower demand by producing less and using less labour and other resources, and thus paying out less incomes (Y) to households. Household consumption will fall and so firms will sell less.

Firms will respond further by producing even less, and thus using even less labour and other resources. Household incomes will fall again. Consumption and hence production will fall again, and so on. There will thus be a multiplied fall in incomes and employment. This is known as the **multiplier effect** and is an example of the 'principle of cumulative causation'.

KEY IDEA 33 TC 14 ***The principle of cumulative causation.*** An initial event can cause an ultimate effect that is much larger. This phenomenon of things building on themselves occurs throughout market economies. It is a fundamental principle in economics and is the fourteenth of our fifteen Threshold Concepts.

The process, however, does not go on forever. Each time household incomes fall, households save less, pay less in taxes and buy fewer imports. In other words, withdrawals fall. When withdrawals have fallen to match the fall in injections, equilibrium will be restored and national income and employment will stop rising. The process can be summarised as follows:

$$W > J \rightarrow Y\downarrow \rightarrow W\downarrow \text{until } J = W$$

Similarly, an initial rise in injections (or fall in withdrawals) will lead to a multiplied rise in national income and employment:

$$J > W \rightarrow Y\uparrow \rightarrow W\uparrow \text{until } J = W$$

Thus equilibrium in the circular flow of income can be at *any* level of output and employment. This was a key insight of Keynes, who argued that mass unemployment could persist unless there were a rise in aggregate demand. This might have to be achieved by the government increasing its expenditure or cutting taxes and this deliberately running a budget deficit.

Showing equilibrium with a Keynesian diagram

We now want to present our simple Keynesian model a little more formally. A number of factors are likely to affect aggregate demand and its components and we will consider some of these in Section 9.6. However, when modelling we tend to simplify matters and so abstract from some of the complex realities of the real world. That is what we are going to do here to gain additional insights into the relationship between aggregate demand and national income. TC 4 p 13

The equilibrium level of national income can be shown on a 'Keynesian' diagram. This plots various elements of the circular flow of income (such as consumption, withdrawals, injections and aggregate demand) against national income (i.e. real GDP). There are two approaches to finding equilibrium: the withdrawals and injections approach; and the income and expenditure approach. Let us examine each in turn.

The withdrawals and injections approach

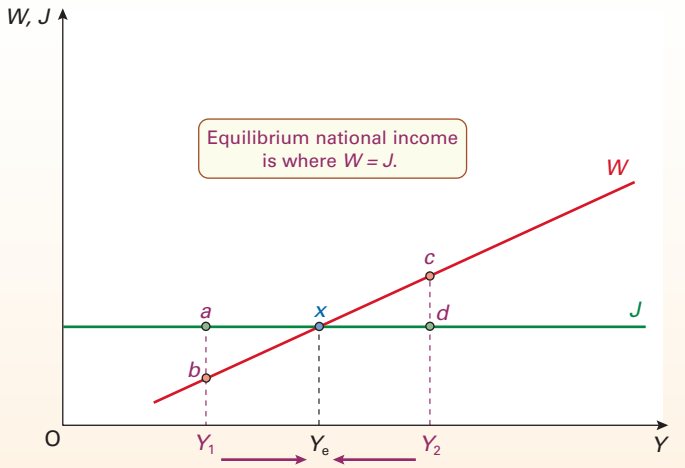
In Figure 9.9, national income (real GDP) (Y) is plotted on the horizontal axis; withdrawals (W) and injections (J) are plotted on the vertical axis.

In constructing Figure 9.9 we assume a *positive* relationship between national income and each of the withdrawals (saving, taxes and imports): in other words, a rise in national income causes a rise in each of the three withdrawals. This gives an upward-sloping withdrawals curve, which, for simplicity, we assume is a straight line.

Definition

Multiplier effect An initial increase in aggregate demand of £xm leads to an eventual rise in national income that is greater than £xm.

Figure 9.9 Equilibrium national income: withdrawals equals injections



The slope of the W curve is given by $\Delta W/\Delta Y$. This is the proportion of a rise in national income that is withdrawn, and is known as the **marginal propensity to withdraw** (**mpw**). Therefore, if one quarter of any rise in national income is withdrawn from the circular flow, the *mpw* is $^1/_4$. With a straight line W 'curve', its slope is constant and so too is the *mpw*.

Pause for thought

Why might withdrawals be negative at very low levels of national income?

Now we turn to the injections line. For simplicity of our model, injections are assumed to be independent of national income. The injections line, therefore, is drawn as a horizontal straight line. This is a reasonable assumption in the short run, where investment and government expenditure plans are likely to be made in advance and where exports depend on foreign incomes, not this country's income. (This does not mean that injections are constant over time: merely that they are constant with respect to national income. If injections rise, the whole line will shift upwards.)

Withdrawals equal injections at point x in the diagram. Equilibrium national income is thus Y_e.

If national income were below this level, say at Y_1, injections would exceed withdrawals (by an amount $a - b$). This additional net expenditure injected into the economy would encourage firms to produce more. This in turn would cause national income to rise. But as people's incomes rose, so they would save more, pay more taxes and buy more imports. In other words, withdrawals would rise. There would be a movement up along the W curve. This process would continue until $W = J$ at point x.

TC 8 p 37

If, on the other hand, national income were initially at Y_2, withdrawals would exceed injections (by an amount $c - d$). This deficiency of demand would cause production and hence national income to fall. As it did so, there would be a movement down along the W curve until again point x was reached.

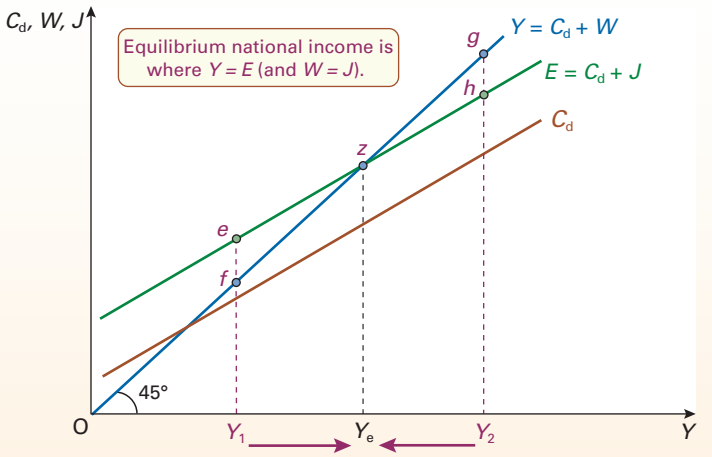
The income and expenditure approach

Another way of showing equilibrium in the Keynesian model is to use a '45° line' diagram (sometimes known as the 'Keynesian cross' diagram).

In Figure 9.10 the 45° line out from the origin plots $C_d + W$ against Y. It is a 45° line because, by definition, $Y = C_d + W$. To understand this, consider what can happen to national income: either it must be spent on domestically produced goods (C_d) or it must be withdrawn from the circular flow – there is nothing else that can happen to it. Thus if Y were £2 trillion, then $C_d + W$ must also be £2 trillion. If you draw a line such that whatever value is plotted on the horizontal axis (Y) is also plotted on the vertical axis ($C_d + W$), the line will be at 45 degrees (assuming that the axes are drawn to the same scale).

The green line plots aggregate demand. In this diagram it is known as the *aggregate expenditure line* (E). It consists of $C_d + J$: in other words, the total spending on the product of domestic firms (see Figure 9.8).

Figure 9.10 Equilibrium national income: real national income equals aggregate expenditure



To show how this line is constructed, consider the brown line. This shows C_d. It is flatter than the 45° line. The reason is that for any given rise in national income, only *part* will be spent on domestic product, while the remainder will be withdrawn: i.e. C_d rises less quickly than Y.

The proportion of the rise in national income which is spent on domestic goods and services is called the **marginal propensity to consume domestically produced goods** (***mpc***$_d$). Therefore, if three quarters of the rise in national income is spent on home-produced items, the mpc_d is $^3/_4$. Since national income must be spent on domestically produced goods (C_d) or withdrawn from the circular flow, the sum of mpc_d and mpw must equal 1. Therefore, if the mpc_d is $^3/_4$ the mpw equals $^1/_4$.

$$mpc_d + mpw = 1$$

The E line consists of $C_d + J$. But we have assumed that J is constant with respect to Y. Thus the E line is simply the C_d line shifted upwards by the amount of J.

If aggregate expenditure exceeded national income, at say Y_1, there would be excess demand in the economy (of $e - f$). In other words, people would be buying more than was currently being produced. Firms would thus find their stocks dwindling and would therefore increase their level of production. In doing so, they would employ more factors of production. National income would thus rise.

As it did so, C_d and hence E would rise. There would be a movement up along the E line. But because not all the extra income would be consumed (i.e. some would be withdrawn), expenditure would rise less quickly than income: the E line is flatter than the Y line. As income rises towards Y_e, the gap between E and Y gets smaller. Once point z is reached, $Y = E$. There is then no further tendency for income to rise.

If national income exceeded aggregate expenditure, at say Y_2, there would be insufficient demand for the goods and services currently being produced ($g - h$). Firms would find their stocks of unsold goods building up. They would thus respond by producing less and employing fewer factors of production. National income would thus fall and go on falling until Y_e was reached.

Pause for thought

a. Why does a − b in Figure 9.9 equal e − f in Figure 9.10?
b. Why does c − d in Figure 9.9 equal g − h in Figure 9.10?

Note that if Y and E, and W and J were plotted on the same diagram, point z (in Figure 9.10) would be vertically above point x (in Figure 9.9).

Definitions

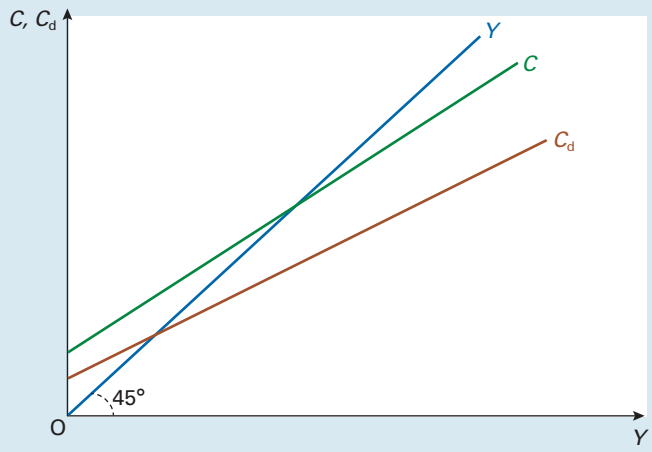
Marginal propensity to withdraw ***(mpw)*** The proportion of an increase in national income that is withdrawn from the circular flow: $\Delta W/\Delta Y$.

Marginal propensity to consume domestically produced goods ***(mpc***$_d$**)** The proportion of a rise in national income that is spent on goods and services produced within the country: $\Delta C_d/\Delta Y$.

BOX 9.2 THE CONSUMPTION FUNCTION

The relationship between consumption and income

Consumption and the consumption of domestic product



At the centre of the simple Keynesian model is the ***Keynesian consumption function***. This proposes that aggregate consumption increases as national income increases. The reason is simple: if people earn more, they can afford to spend more.

We can show the consumption function graphically on the 45° line diagram. Note that here we are talking about *total* consumption, which includes the consumption of imports as well as domestically produced goods and services.

To simplify the analysis, the consumption–income relationship is drawn as a straight line. At very low levels of income, the consumption function will lie above the 45° line. When a nation is very poor, most people may be forced to spend more than they earn merely to survive. They usually do this by borrowing or drawing on savings. Above a certain level of income, however, the consumption function will lie below the 45° line. People will spend less than they earn. The remainder will go on saving and taxes.

The higher the level of national income, the smaller the proportion that will be consumed: people can afford to save proportionately more, and will have to pay proportionately more in taxes. In other words, the ***average propensity to consume (apc)*** – the ratio C/Y – falls as national income rises. It follows that the slope of the consumption function is less than that of the 45° line.

The slope of the consumption function is given by the *marginal propensity to consume (mpc)*. This is the

Recap

1. In the simple Keynesian model, equilibrium national income is where withdrawals equal injections, and where national income equals the total expenditure on domestic products: where $W = J$ and where $Y = E$.
2. The relationships between national income and the various components of the circular flow of income can be shown on a diagram, where national income is plotted on the horizontal axis and the various components of the circular flow are plotted on the vertical axis.
3. Equilibrium national income can be shown on this diagram, either at the point where the W and J lines cross or where the E line crosses the 45° line (Y).

9.5 THE MULTIPLIER

What will be the effect on output of a change in aggregate expenditure?

In a demand-driven model of the economy, when injections rise (or withdrawals fall) national income will rise. But by how much? The answer is that there will be a *multiplied* rise in income: i.e. national income will rise by *more* than the rise in injections (or fall in withdrawals). The size of the **multiplier** is given by the letter k, where:

TC4 p 13

EXPLORING THE THEORY

proportion of any increase in income that goes on consumption: i.e. $\Delta C/\Delta Y$. In our diagram, the *mpc* is shown as the propensity to consume out of additional *national* income. Because the consumption function is a straight line, it has a *constant* slope and hence the *mpc* is also constant.

The *mpc* is sometimes defined as the proportion of a rise in *disposable* national income that goes on consumption, where disposable income is income after the receipt of cash benefits and the payment of income tax and social insurance contributions. By modelling consumption as a function of *gross* national income (i.e. before taxes and benefits), the Keynesian model developed in Section 9.4 is simplified without affecting its general conclusions for the determination of national income.

Movement along and shifts in the consumption function

The effect on consumption of a change in national income is shown by a movement *along* the consumption function. Of course, people's current incomes are not the only determinants of the amount they consume. There are several other determinants. These include expected future incomes, consumer confidence, economic uncertainty and household wealth and financial well-being.

A change in any of the 'other determinants' of consumption is shown by a *shift* in the consumption function. We will explore these 'other determinants' in more detail in Section 9.5 when reflecting further on national income determination and economic volatility.

Consumption of domestic product (C_d)

The parts of consumption that go on imports and indirect taxes constitute *withdrawals* from the circular flow of income and thus do not contribute to aggregate demand. That part of consumption that *does* contribute to aggregate demand is known as the *consumption of domestic product* or of domestically produced goods and services (C_d). Therefore, the C_d function lies below the *C* function. The gap between them constitutes imports of consumer goods and expenditure taxes, such as VAT and excise duties.

1. *What factors could influence an individual's propensity to consume additional income?*
2. *What determinants of consumption could lead to a vertical upward shift of the consumption function?*

Using news articles, discuss the actual relationship between income and consumption and identify factors other than income which have affected consumer spending.

Definitions

Keynesian consumption function The relationship between consumption and income. It can be expressed algebraically or graphically.

Average propensity to consume The proportion of income that is spent (C/Y).

$$k = \Delta Y/\Delta J$$

Thus if injections rose by £10 million (ΔJ) and, as a result, national income rose by £30 million (ΔY), the multiplier would be 3.

But what determines the size of the rise in income (ΔY)? In other words, what determines the size of the multiplier? This can be shown graphically using either the withdrawals and injections approach or the income and expenditure approach from the previous section. (You may omit one, if you choose.)

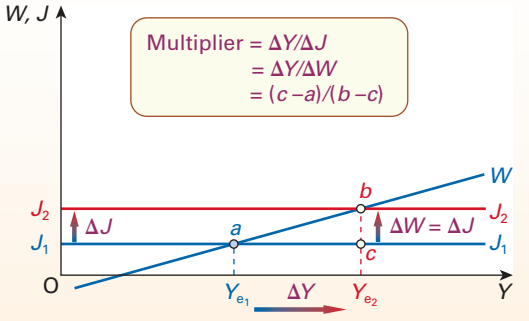
TC 4 p 13

The withdrawals and injections approach

Assume that injections rise from J_1 to J_2 in Figure 9.11. Equilibrium will move from point *a* to point *b*. Income will thus rise from Y_{e1} to Y_{e2}. But this rise in income (ΔY) is bigger than the rise in injections (ΔJ) that caused it. This is the multiplier effect. It is given by $(c - a)/(b - c)$ (i.e. $\Delta Y/\Delta J$).

It can be seen that the size of the multiplier depends on the *slope of the W curve*: the *marginal propensity to withdraw* (*mpw*). The flatter the curve and the smaller the *mpw*, the bigger will be the multiplier and so the bigger will be the rise in national income from any given rise in injections. The point here is that the less is withdrawn each time money circulates, the more

Figure 9.11 The multiplier: a shift in injections



will be re-circulated and hence the bigger the rise in national income will be.

The size of the multiplier thus varies inversely with the size of the *mpw*. The bigger the *mpw*, the smaller the multiplier; the smaller the *mpw*, the bigger the multiplier. In fact the **multiplier formula** is simply the inverse of the *mpw*:

$$k = \frac{1}{mpw}$$

Thus if the *mpw* were $^1/_4$, the multiplier would be 4. So if *J* increased by £100 million, *Y* would increase by £400 million.

To understand why, consider what must happen to withdrawals. Injections have risen by £100 million, thus withdrawals must rise by £100 million to restore equilibrium ($J = W$). But with an *mpw* of $^1/_4$, this £100 million rise in withdrawals must be one-quarter of the rise in national income that has resulted from the extra injections. Thus *Y* must rise by £400 million.

An alternative formula uses the concept of the marginal propensity to consume domestically produced goods (mpc_d) that we introduced earlier. This, as we saw, is the proportion of a rise in national income that is spent on domestically produced goods, and thus is not withdrawn. Thus if a quarter of a rise in national income is withdrawn, the remaining three-quarters will re-circulate as C_d. Since $mpw = 1 - mpc_d$, an alternative formula for the multiplier is:

$$k = 1/(1 - mpc_d)$$

But why is the multiplier given by the formula $1/mpw$? This can be illustrated by referring to Figure 9.11. The *mpw* is the slope of the *W* line. In the diagram this is given by the amount $(b - c)/(c - a)$. The multiplier is defined as $\Delta Y/\Delta J$. In the diagram this is the amount $(c - a)/(b - c)$. But this is merely the inverse of the *mpw*. Thus the multiplier equals $1/mpw$.[3]

The income and expenditure approach

The multiplier can also be demonstrated using the income/expenditure approach. Assume in Figure 9.12 that the aggregate expenditure function shifts to E_2. This could be due either to a rise in one or more of the three injections, or to a rise in the consumption of domestically produced goods (and hence a fall in withdrawals). Equilibrium national income will rise from Y_{e1} to Y_{e2}.

What is the size of the multiplier? The initial rise in expenditure was $b - a$. The resulting rise in income is $c - a$. The multiplier is thus $(c - a)/(b - a)$.

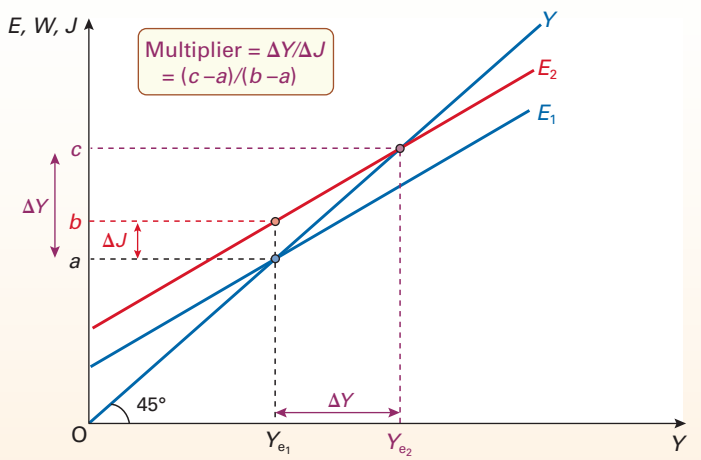
Definitions

Multiplier The number of times by which a rise in national income (ΔY) exceeds the rise in injections (ΔJ) that caused it: $k = \Delta Y/\Delta J$.

Multiplier formula The formula for the multiplier is: $k = 1/mpw$ or $1/(1 - mpc_d)$.

[3] In some elementary textbooks, the formula for the multiplier is given as $1/mps$ (where *mps* is the marginal propensity to save: the proportion of a rise in income saved). The reason for this is that it is assumed (for simplicity) that there is only one withdrawal, namely saving, and only one injection, namely investment. As soon as this assumption is dropped, $1/mps$ becomes the wrong formula.

Figure 9.12 The multiplier: a shift in the expenditure function



Pause for thought

Why might a country have a steep E *line, and hence a high value for the multiplier?*

The multiplier: a numerical illustration

The multiplier effect does not work instantaneously. When there is an increase in injections, whether investment, government expenditure or exports, it takes time before this brings about the full multiplied rise in national income.

Consider the following example. Let us assume for simplicity that the mpw is $^1/_2$. This will give an mpc_d of $^1/_2$ also. Let us also assume that investment (an injection) rises by £160 million and stays at the new higher level. Table 9.2 shows what will happen.

As firms purchase more machines and construct more factories, the incomes of those who produce machines and those who work in the construction industry will increase by £160 million. When this extra income is received by households, whether as wages or profits, half will be withdrawn ($mpw = {}^1/_2$) and half will be spent on the goods and services of domestic firms. This increase in consumption thus generates additional incomes for firms of £80 million over and above the initial £160 million (which is still being generated in each time period).

When this additional £80 million of incomes is received by households (round 2), again half will be withdrawn and half will go on consumption of domestic product. This increases national income by a further £40 million (round 3). And so each time we go around the circular flow of income, national income increases, but by only half as much as the previous time ($mpc_d = {}^1/_2$).

If we add up the additional income generated in each round (assuming the process goes on indefinitely), the total will be £320 million: twice the rise in injections. The multiplier is 2.

Table 9.2 The multiplier 'round'

Round	ΔJ (£m)	ΔY (£m)	ΔC_d (£m)	ΔW (£m)
1	160	160	80	80
2	–	80	40	40
3	–	40	20	20
4	–	20	10	10
5	–	10	5	5
6	–	5	...	...
.	.	.	.	.
$1 \rightarrow \infty$	320	160	160	160

The bigger the mpc_d (and hence the smaller the mpw), the more expenditure will rise each time national income rises, and hence the bigger the multiplier will be.

The multiplier and the full-employment level of national income

The simple Keynesian theory assumes that there is a maximum level of national income which can be obtained at any one time. If the equilibrium level of income is at this level, there will be no deficiency of aggregate demand. This level of income is referred to as the **full-employment level of national income**. In practice, although at this level of national income there would be no cyclical unemployment, there would still be some unemployment for other reasons. We explore other types of unemployment in Section 12.4.

Recessionary gap

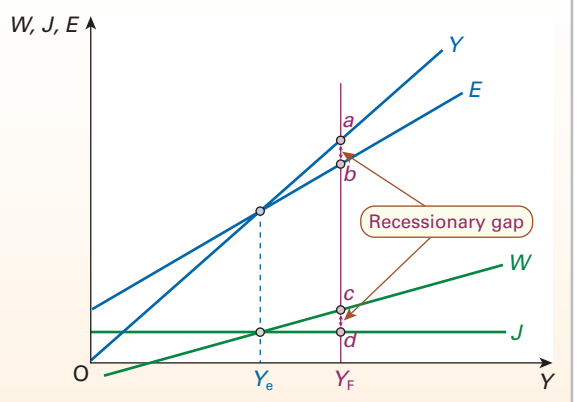
Unfortunately, the equilibrium level of national income (Y_e) could fall *below* the full-employment level (Y_F) resulting in a **recessionary gap** (also known as a **deflationary gap**) and **demand-deficient unemployment**. This is illustrated in Figure 9.13.

The full-employment level of national income (Y_F) is represented by the vertical line. The equilibrium level of national income is Y_e, where $W = J$ and $Y = E$. The deflationary gap is $a - b$: namely, the amount that the E line is below the 45° line at the full-employment level of income (Y_F). It is also $c - d$: the amount that injections fall short of withdrawals at the full-employment level of income.

Note that the size of the recessionary gap is *less* than the amount by which Y_e falls short of Y_F. This provides another illustration of the multiplier. If injections were raised by $c - d$, income would rise by $Y_F - Y_e$. The multiplier is thus given by:

TC 14 p 249

Figure 9.13 The recessionary gap



$$\frac{Y_F - Y_e}{c - d}$$

In this simple Keynesian model, then, the cure for demand-deficient unemployment is to close the deflationary gap. To close the deflationary gap an increase in aggregate expenditure is needed. This could be achieved by an expansionary *fiscal* policy of increasing government expenditure and/or lowering taxes, or by an expansionary *monetary* policy of reducing interest rates and increasing the amount of money in the economy, thereby encouraging extra consumption and investment. We examine such policies in Chapter 13.

Inflationary gap

The equilibrium level could lie *above* the full-employment level. This situation involves an **inflationary gap**. This is the amount by which aggregate expenditure exceeds national income or injections exceed withdrawals at the full-employment level of national income. This is illustrated by the gaps $e - f$ and $g - h$ in Figure 9.14.

The problem is that Y_F represents a real ceiling to output. Real national income cannot expand beyond this point, other than for a very short time where the factors of production are used very intensively, for example with labour working overtime. This, however, is not sustainable and so results in inflation.[4] Inflation resulting from excess demand is known as **demand-pull inflation**.

To eliminate this inflation, the inflationary gap must be closed by either raising withdrawals or lowering injections, or some combination of the two, until Y_e equals Y_F. This could involve either a contractionary *fiscal* policy of lowering government expenditure and/or raising taxes, or a contractionary *monetary* policy of raising interest rates and reducing the amount of money in the economy.

Even if the government does not actively pursue a deflationary policy, the inflationary gap may still close *automatically*. The mechanisms by which this could occur would move the *E* line down and the *J* line up and/or the *W* line up. These include the following:

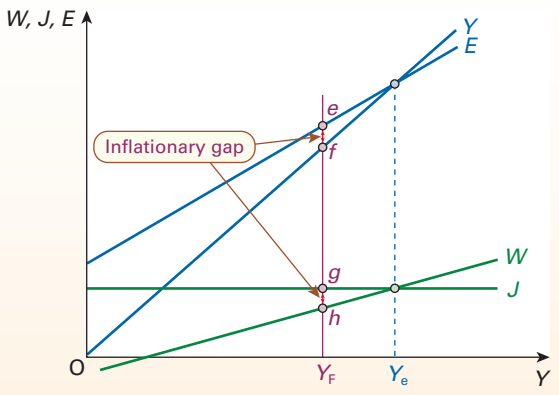
- When prices rise, people's wages may not rise in line, at least not in the short run. Consumers could therefore see a cut in real incomes and thus spend less.
- Higher domestic prices will lead to fewer exports being sold (lower injections) and more imports being bought (higher withdrawals) in preference to the now more expensive home-produced goods.
- Higher prices reduce the real value of people's savings. They may therefore save more to compensate for this.
- Higher prices increase the demand for money. The average amount of money that people and firms would need to hold for spending purposes is greater. In the absence of an increase in the money supply, the shortage of money drives up interest rates. This reduces investment and encourages saving. We examine the money market in Chapter 11.

Pause for thought

How would each of these possible mechanisms affect the J and W lines in the injections/withdrawals diagram?

The problem for policy makers is that these automatic effects may take some time to take effect and, even then, may be difficult to predict.

Figure 9.14 The inflationary gap



[4]Note that the horizontal axis in the 45° line diagram represents *real* national income. If incomes were to rise by, say, 10 per cent but prices also rose by 10 per cent, real income would not have risen at all. People could not buy any more than before. In such a case, there will have been no rightward movement along the horizontal axis.

Definitions

Full-employment level of national income The level of national income (real GDP) at which there is no deficiency of demand.

Recessionary or deflationary gap The shortfall of national expenditure below national income (and injections).

Demand-deficient (or cyclical) unemployment Unemployment caused by a lack of aggregate demand. Firms, unable to sell all they produce, lay off workers.

Inflationary gap The excess of national expenditure over income (and injections over withdrawals) at the full-employment level of national income.

Demand-pull inflation Inflation caused by persistent rises in aggregate demand beyond the ability of the economy to supply.

Inflation and the multiplier

The simple Keynesian model implies that up to the full-employment level of national income (Y_F), output and employment can increase with no rise in prices at all. There is no inflation because the economy's deflationary gap is being closed. Hence, increases in aggregate expenditure set in motion the multiplier process and the result is a multiplied increase in real national income.

However, according to the model, beyond Y_F further rises in aggregate demand are entirely reflected in higher prices. An inflationary gap opens. Hence, the additional demand no longer generates an increase in output. There is no increase in real national income because the real value of expenditure has not increased. Additional sums of spending on domestically produced goods and service are purely *nominal* reflecting only higher prices. The volume of purchases and, hence, the level of output is unchanged.

TC 5 p 17

In this simple model, therefore, resource constraints become effective at the full-employment national income level. Any further increases in aggregate demand generate only inflation. The multiplier process ceases to operate. We can use the income/expenditure approach to illustrate these ideas as shown in Figure 9.15.

Assume that the economy is initially at Y_{e1} where E_1 crosses the 45° line. Now let us assume that there is a rise in aggregate demand. The aggregate expenditure function shifts to E_2 resulting in a full multiplied rise in real income. Therefore, equilibrium national income rises to Y_F.

Consider now a further rise in aggregate demand which causes the aggregate expenditure function to shift to E_3. An inflationary gap opens up, illustrated by the gap $e - f$. This time the increase in demand will be reflected only in higher prices with no increase in output. Equilibrium *real* income will be unchanged. If there is no compensating increase in money supply, the E line will tend to fall back to E_2. The means by which this happens are the automatic mechanisms identified above. Alternatively fiscal or money policy may be used to shift the E curve back to E_2.

TC 5 p 17

The simple Keynesian model suggests, therefore, that inflation only occurs Y_F when is reached and the E line rises above E_2 in Figure 9.15. In reality, inflation may begin to occur *before* the full-employment level of income is reached. For example, not all firms operate with the same degree of slack. Therefore, some firms may respond to a general rise in aggregate demand by taking up slack and hence increasing output, while other firms, having little or no slack, respond by raising prices.

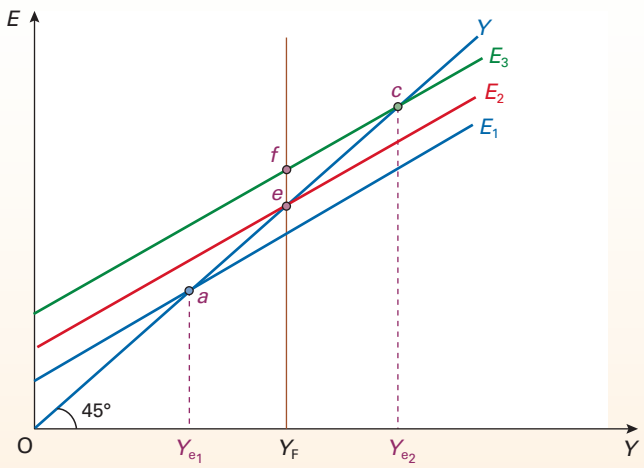
The simple Keynesian model of the economy shows how the level of aggregate demand determines the level of national income. In this model, therefore, the causes of cyclical fluctuations in the economy are fluctuations in aggregate demand: i.e. fluctuations in the consumption of domestic product (C_d), withdrawals (W) and injections (J).

Their effects on output and employment are then amplified by the multiplier process. The amplification of the peaks and troughs of the business cycle is greater the larger is the multiplier.

TC 14 p 249

But what might be the potential sources of this volatility in aggregate demand? We discuss this in the final section of the chapter.

Figure 9.15 Allowing for inflation: the income/expenditure approach



Recap

1. If injections rise (or withdrawals fall), there will be a multiplied rise in national income. The multiplier is defined as $\Delta Y/\Delta J$. Thus if a £10 million rise in injections led to a £50 million rise in national income, the multiplier would be 5.
2. The size of the multiplier depends on the marginal propensity to withdraw (*mpw*). The smaller the *mpw,* the less will be withdrawn each time incomes are generated round the circular flow, and thus the more will go round again as *additional* demand for domestic product.
3. The multiplier formula is $1/mpw$ or $1/(1 - mpc_d)$.
4. The full-employment output level acts as a constraint on the multiplier process. Further increases in aggregate demand are inflationary and the multiplier process ceases to operate. In reality, inflationary effects may begin to occur before the full-employment level and so the impact on output from increases in aggregate demand may be less than implied by the simple Keynesian model.

9.6 THE VOLATILITY OF PRIVATE-SECTOR SPENDING

What causes spending to fluctuate?

The circular flow and simple Keynesian models show that changes in spending by the interdependent set of groups that comprise the economy (households, firms, government, financial institutions and the overseas sector) can have significant short-run effects on the economy. Hence, *short-run* changes or fluctuations in aggregate demand create fluctuations in levels of output and employment. In the long run, it is generally thought that changes in aggregate demand will have much less impact on output and employment and much more effect on prices.

TC 3 p 10

In order to analyse what drives *changes* in the level of aggregate demand, we need to consider further some of the possible influences on the spending decisions of the purchasers of goods and services. Here we focus predominantly on fluctuations in domestic private-sector spending. We will consider government spending in Chapter 13 and the interaction between the national and overseas economies in Chapters 14 and 15.

Consumption cycles

When analysing the role played by the private sector as a source of economic volatility, it makes sense to begin by looking at consumer spending. As we saw in Table 9.1, in most countries the largest expenditure component of aggregate demand by value is household consumption (C). In the UK, for example, it frequently accounts for over 60 per cent of national income. This means that even small fluctuations in consumer expenditure can be significant for aggregate demand.

As Figure 9.16 shows, annual rates of economic growth mirror closely those in household consumption. In comparison to consumption, investment levels fluctuate significantly more. We will return to this shortly. The key point though is that the business cycle tends to reflect the consumption cycle.

TC 3 p 10

But what factors determine consumption? Which are most important? How do changes in consumption then affect output and other macroeconomic variables? Do these effects then feed back to influence further the level of consumption? Is the determination of consumption predictable?

The Keynesian consumption function (see Box 9.2) suggests that current levels of national income determine aggregate consumption levels. But, clearly there are other factors influencing people's spending decisions.

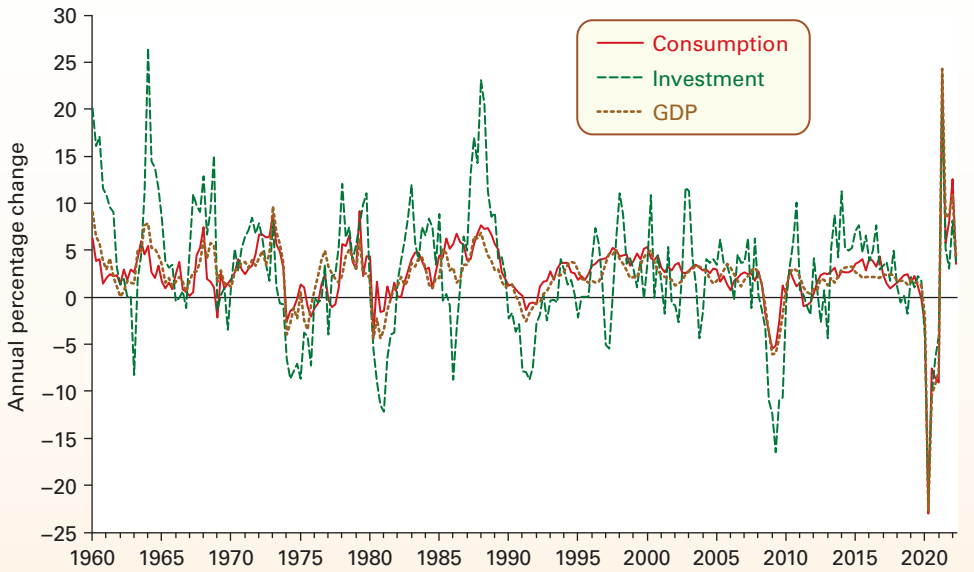
Taxation. The higher the level of direct taxes (income tax and social insurance contributions), the less people will have left to spend out of their gross income: consumption depends on *disposable income*. Disposable income also rises when government increases cash benefits.

Expected future incomes. Many people take into account both current and expected future incomes when planning their current and future consumption. In other words, they are forward looking. You might have a relatively low income when you graduate, but can expect (you hope!) to earn much more in the future.

TC 13 p 195

You are thus willing to take on more debts now in order to support your consumption, not only as a student but shortly afterwards as well, anticipating that you will be able to pay back these loans later. It is similar with people taking out a mortgage to buy a house. They might struggle to pay the interest at first, but hope that this will become easier over time.

Figure 9.16 Annual growth of UK consumption, investment and output



Notes: Annual growth rates are calculated using constant-price data; investment is the rate of growth of gross fixed capital formation (whole economy)

Source: Based on series *KGZ7*, *KG7T* and *IHYR* (ONS, 2022)

The financial system (such as banks and building societies) is fundamental to the **smoothing of consumption** by households. You can borrow when your income is low and pay back the loans later on when your income is higher. Therefore, the financial system provides households with greater flexibility as to when to spend their expected future incomes.

Pause for thought

If a consumption-smoothing household expects a large payment to be made to their bank account later in a few months' time, what impact will this have on their spending now?

The smoothing of consumption can therefore play a role in reducing the peaks or troughs of the consumption and business cycles. However, expectations of future incomes do get revised. Large revisions can have significant effects for current spending levels. If, for example, there is a general belief among people that future incomes will grow much less quickly than was previously expected, as was the case in many countries in the late 2000s following the financial crisis, then this can significantly dampen today's spending levels.

The financial system and the availability of credit. The financial sector provides households with both longer-term loans and also short-term credit. Short-term credit enables transactions to take place by bridging short-term gaps between income and expenditure flows. This helps to explain why short-term rates of change in household spending, such as those from one quarter of the year to the next, are generally less variable than those in disposable income.

KI 23 p 173

However, the growth in consumption is affected if the ability and willingness of financial institutions to provide credit changes. The global financial crises of the second half of the 2000s saw credit criteria tighten dramatically. A tightening of credit practices, such as reducing overdraft facilities or reducing income multiples (the size of loans made available relative to household incomes), weakens consumption growth. This is because the growth of consumption becomes more dependent on the growth of current incomes. Consequently, there is a growth in the number of **credit-constrained households**.

In contrast, a relaxation of lending practices, as seen in many countries during the 1980s, can strengthen consumption growth. This allows households to borrow more readily against future incomes: there are fewer credit-constrained households.

Pause for thought

What impact do credit constraints have on the short-term relationship between consumption and income?

Interest rates and cash flow effects. Changes in interest rates can affect household spending. They generate cash flow effects by affecting the interest receipts of

savers (creditors) and interest payments by borrowers (debtors). With financialisation and the general increase seen in many countries in people's levels of indebtedness to financial institutions, changes in interest rates can have a significant impact on the costs to households of 'servicing' their loans.

Debt-servicing costs are the costs incurred in repaying loans and the interest payments on them. Where the rate of interest on debt is variable – a variable-rate loan – changes in interest rates affect the cost of servicing the debt. These effects can be especially important for mortgages. In the UK, where a large proportion of mortgage-payers are on a variable mortgage rate, changes in mortgage rates can have a sizeable impact on their debt-servicing costs. This then increases the proportion of current income that a household needs to set aside to pay the mortgage and reduces the proportion that could otherwise have been used for consumption.

Pause for thought

How might the marginal propensity to consume of debtors differ from that of creditors?

Wealth and household sector balance sheets. By borrowing and saving, households accumulate a stock of financial liabilities (debts), financial assets (savings) and physical assets (mainly property). The household sector's *financial* balance sheet details the sector's holding of financial assets and liabilities, while its *capital* balance sheet details its physical assets. The balance of financial assets over liabilities is the household sector's net financial wealth. The household sector's **net worth** is the sum of its net financial wealth and its physical wealth.

Table 9.3 shows the summary balance sheet of the UK household sector in 1995 and 2021. At the end of 2021 the sector had a stock of net worth estimated at £11.77 trillion (7.9 times annual household disposable income or 5.1 times GDP) compared with £2.7 trillion (4.7 times annual household disposable income or 3.1 times GDP) at the end of 1995 – an increase of 340 per cent. This, of course, is a nominal increase, not a real increase, as part of it merely reflects the rise in asset prices.

Changes to the household sector balance sheets affect the sector's financial well-being (or its financial distress). These changes may relate to particular components, for example the level of liquidity provided by financial assets on the financial balance sheet, or to broader aggregates such as net financial wealth or net worth.

The effects on spending from a sector's balance sheets are known as **balance sheet effects**. These effects apply not only to households but to all sectors of the economy. In the context of households, a deterioration in financial well-being would be expected to dampen consumer spending. For example, a declining net worth-to-income ratio, perhaps caused by falling house prices or falling share prices, could lead to the household sector increasing its savings or repaying some of its outstanding debts. On the other hand, improvements to financial well-being can raise consumption levels.

Consumer confidence. Consumer confidence relates to the sentiment, emotion, or anxiety of consumers. At its most simple, consumer (and business)

Definitions

Consumption smoothing The act by households of smoothing their levels of consumption over time despite facing volatile incomes.

Credit-constrained households Households that are limited in their ability to borrow against expected future incomes.

Debt-servicing costs The costs incurred when repaying debt, including debt interest payments.

Table 9.3 Summary balance sheet of household sector, 1995 and 2021

	1995			2021		
	£ billions	% of disposable income	% of GDP	£ billions	% of disposable income	% of GDP
Financial assets	2065.1	365.4	242.0	7430.6	496.2	320.7
Financial liabilities	541.2	95.8	63.4	2044.1	136.5	88.2
Net financial wealth	**1523.9**	**269.6**	**178.6**	**5386.6**	**359.7**	**232.5**
Non-financial assets	1150.5	203.6	134.8	6386.1	426.4	275.6
Net worth	**2674.3**	**473.2**	**313.4**	**11 772.7**	**786.1**	**508.1**

Source: Based on data from *The UK National balance sheet estimates: 2021* and series *YBHA* and *HABN* (National Statistics, 2022)

Pause for thought

In recent years, there has been an increase in the use of individual voluntary arrangements (IVAs) whereby people who have got into financial trouble can arrange a repayment schedule with their creditors. These arrangements can involve some of the debt being written off. What is the effect likely to be on borrowing? Is there a moral hazard here (see page 83)?

confidence surveys (see Box 9.3) are trying to assess feelings of optimism or pessimism, particularly in relation to the economy or financial well-being. These surveys aim to shed light on spending intentions and hence the short-term prospects for consumption. A rise in confidence would be expected to lead to a rise in consumption, while a fall would lead to a fall in consumption.

Uncertainty. If people are uncertain about their future income prospects, or fear unemployment, they are likely to be cautious in their spending. Heightened uncertainty tends to erode confidence and hence increase saving. People become more prudent.

KI 15 p 80

Saving undertaken by people (or businesses) to guard again uncertainty, such as in future income levels, is known as **precautionary** or **buffer-stock saving**. The resulting stock of savings acts as a buffer that can, at some future point, be readily converted into a given amount of cash. This of course means foregoing an amount of consumption now.

TC 6 p 24

An increase in economic uncertainly creates an incentive for people to increase their level of buffer-stock savings. Consumption levels fall. On the other hand, with less uncertainty people may feel more confident in holding a smaller buffer stock. The result is greater spending. They become less prudent.

Pause for thought

How does uncertainty affect the relative importance of 'prudence' and 'impatience' in determining people's current spending plans?

TC 11 p 61

Expectations of future prices. If people expect prices to rise, they tend to buy durable goods such as furniture and cars before this happens. Conversely, if people expect prices to fall, they may wait. This has been a problem in Japan for many years, where periods of falling prices (deflation) led many consumers to hold back on spending, thereby weakening aggregate demand and hence economic growth.

The distribution of income. Poorer households will typically spend more than richer ones out of any additional income they receive. They have a higher marginal propensity to consume (*mpc*) than the rich, with very little left over to save. A redistribution of national income from the poor to the rich will therefore tend to reduce the total level of consumption in the economy.

Tastes and attitudes. If people have a 'buy now, pay later' mentality, or a craving for consumer goods, they are likely to have a higher level of consumption than if their tastes are more frugal. The more 'consumerist' and materialistic a nation becomes, facilitated by its financial system, the higher its consumption will be for any given level of income.

The age of durables. If people's car, carpets, clothes, etc. are getting old, they will tend to have a high level of 'replacement' consumption, particularly after a recession when they had cut back on their consumption of durables. Conversely, as the economy reaches the peak of the boom, people are likely to spend less on durables as they have probably already bought the items they want. This behaviour therefore amplifies the consumption cycle.

Instability of investment

Investment (*I*) is the most volatile expenditure component of aggregate demand. This is evident from Figure 9.16 (see page 259). But why is this the case? What drives this volatility?

The investment accelerator

When an economy begins to recover from a recession, investment can rise very rapidly. When the growth of the economy slows down, however, investment can fall dramatically, and during a recession it can all but disappear. Since investment is an injection into the circular flow of income, these changes in investment will cause multiplied changes in income and thus heighten a boom or deepen a recession.

TC 14 p 249

The theory that relates investment to *changes* in national income is called the **accelerator theory**. The term 'accelerator' is used because a relatively modest rise in national income can cause a much larger percentage rise in investment.

Definitions

Net worth The balance of net financial wealth and non-financial assets.

Balance sheet effects The effects on spending behaviour, such as consumer spending, that arise from changes in the composition or value of net worth.

Precautionary or buffer-stock saving Saving in response to uncertainty, for example uncertainty of future income.

When there is no change in income and hence no change in consumption, the only investment needed is a relatively small amount of replacement investment for machines that are wearing out or have become obsolete. When income and consumption increase, however, there will have to be *new* investment in order to increase production capacity. This is called **induced investment** (I_i). Once this has taken place, investment will fall back to mere replacement investment (I_r) unless there is a further rise in income and consumption.

Thus induced investment depends on *changes* in national income (ΔY):

$$I_i = \alpha \Delta Y$$

where α is the amount by which induced investment depends on changes in national income, and is known as the **accelerator coefficient**. Thus if a £1 billion *rise* in national income caused the *level* of induced investment to be £2 billion, the accelerator coefficient would be 2.

The size of α depends on the economy's **marginal capital-output ratio** ($\Delta K/\Delta Y$). If an increase in the country's capital stock of £2 million (i.e. an investment of £2 million) is required to produce £1 million extra national output, the marginal capital/output ratio would be 2. Other things being equal, the accelerator coefficient and the marginal capital/output ratio will therefore be the same.

Case Study 9.12 on the student website uses a numerical example to illustrate some of the important features of the accelerator.

In practice, the size of the accelerator effect is likely to be difficult to predict. Firms will generally have some spare capacity and/or carry stocks which means that they may be in a position to meet extra demand without having to invest. Machines do not as a rule suddenly wear out. A firm could thus delay replacing machines and keep the old ones for a bit longer if it was uncertain about its future level of demand.

Nevertheless, the accelerator effect still appears to exist (see Figure 9.16 and also Case Study 9.13 on the student website) with firms' investment spending responding to changes in aggregate demand.

The interaction of the multiplier and accelerator

TC 14 p 249

Once an accelerator effect takes place, it is then amplified by the multiplier. Similarly, an initial multiplier effect of, say, a rise in consumption can lead to an accelerator effect. The accelerator and multiplier interact. This amplifies the ups and downs of the business cycle.

TC 3 p 10

For example, if there is a rise in government purchases (G), this will lead to a multiplied rise in national income. But this *rise* in national income will set off an accelerator effect: firms will respond to the rise in income and the resulting rise in consumer demand by investing more. But this rise in investment constitutes a further rise in injections and thus will lead to a second multiplied rise in income. If this rise in income is larger than the first, there will then be a second rise in investment (the accelerator), which in turn will cause a third rise in income (the multiplier). And so the process continues indefinitely.

But does this lead to an exploding rise in national income? Will a single rise in injections cause national income to go on rising for ever? The answer is no, for two reasons. The first is that national income, in real terms, cannot go on rising faster than the growth in potential output. It will bump up against the ceiling of full employment, whether of labour or of other resources.

A second reason is that, if investment is to go on rising, it is not enough that national income should merely go on *rising*: instead, national income must *rise faster and faster*. Once the growth in national income slows down, investment will begin to fall, and then the whole process will be reversed. A fall in investment will lead to a fall in national income, which will lead to a massive fall in investment. The multiplier/accelerator interaction is shown more formally in Case Study 9.14 on the student website.

Other determinants of investment

There are other major determinants of investment. These too may contribute the volatility to of investment spending.

Expectations of future market conditions. Since investment is made in order to produce output for the future, investment must depend on firms' expectations about future market conditions. But, this raises important questions about how people form their expectations. After all, the future cannot be predicted with accuracy. To what extent do firms, for instance, use current conditions or those of the recent past as a short-cut in their decision-making process?

TC 13 p 195

Definitions

Accelerator theory The *level* of investment depends on the *rate of change* of national income, and as a result tends to be subject to substantial fluctuations.

Induced investment Investment that firms make to enable them to meet extra consumer demand.

Accelerator coefficient The level of induced investment as a proportion of a rise in national income: $\alpha = I_i/\Delta Y$.

Marginal capital-output ratio The amount of extra capital (in money terms) required to produce a £1 increase in national output. Since $I_i = \Delta K$, the marginal capital/output ratio $\Delta K/\Delta Y$ equals the accelerator coefficient (α).

If current economic growth helps determine firms' expectations of future growth, it increases the possibility of *bandwagon effects* affecting investment levels. Hence once the economy starts expanding, expectations become buoyant and firms increase investment which, through the multiplier and accelerator, boosts economic growth and further increases confidence. Likewise in a recession, a mood of pessimism may set in and firms cut investment.

TC 14 p 249

Uncertainty. Since future market conditions cannot be predicted with accuracy, investment is *risky*. Greater uncertainty about these conditions will make firms more cautious to invest. In the same way that households look to self-insure against uncertainty, firms are less likely to proceed with investment projects when uncertainty is high. Hence, greater uncertainty negatively affects levels of investment and aggregate demand. Uncertainty over the effects of Brexit led to a fall in investment following the 2016 referendum.

KI 15 p 80

Business confidence. Confidence in the context of business relates to the sentiment or anxiety of firms about market conditions. Investment depends crucially on business confidence. The scale and cost of capital projects by firms can be very large indeed and their effects often irreversible. Hence, increased pessimism among businesses may reduce investment quite significantly.

The availability of finance. Investment often requires financing and this can involve very significant amounts of money. Firms may seek finance from banks, or perhaps issue debt instruments, such as bonds, or issue new shares. Therefore, difficulties in raising finance, such as were seen in the late 2000s and into the early 2010s, can limit investment.

The rate of interest. The higher the rate of interest, the more expensive it will be for firms to finance investment, and hence the less profitable the investment will be. Economists keenly debate just how responsive total investment in the economy is to changes in interest rates.

The cost and efficiency of capital equipment. If the cost of capital equipment goes down or machines become more efficient, the return on investment will increase. Firms will invest more. Technological progress is an important determinant here.

Pause for thought

To what extent are the 'other determinants' of investment also likely to be determinants of household consumption?

The role of the financial sector

One sector that plays a crucial role in affecting economic activity is the financial sector. We have already seen several ways in which it can affect spending. Sometimes it can dampen expenditure cycles by helping people and businesses to smooth their spending through the provision of short-term credit. On other occasions, it appears to create volatility. This might be because of changes in the availability or cost of credit or through financial institutions' impact on the balance sheets of different sectors of the economy.

We will focus in detail on the markets, institutions and products and services which comprise the financial sector in Chapter 11. Yet its economic significance means that it will be referred to frequently in subsequent chapters.

The financial crisis of the late 2000s has resulted in more analysis of the role played by the financial sector in shaping or perhaps even causing the business cycle.

The financial sector as a source of volatility. Some economists argue that the financial sector is a major *source* of economic volatility. Some go as far as to say that the behaviour of financial institutions through their lending and investments generates unsustainable economic growth which inevitably ends with an economic downturn. This behaviour leads to deteriorating financial well-being.

The 2007–9 financial crisis is for some economists evidence of the phenomenon of a **balance sheet recession**: an economic downturn caused by financially distressed sectors of the economy. The term is applied to sectors which become overly burdened by their debts. Policy makers therefore keep a close eye on various indicators of financial well-being across different sectors, including the household sector. Financial distress can be an early warning sign of a possible impending crisis and recession.

KI 31 p 206

Pause for thought

Can economies experience balance sheet booms?

The financial sector as a magnifier of volatility. Other economists argue that the financial sector *amplifies* economic shocks. The argument here is not that financial institutions are the source of fluctuations in economic growth but rather that they magnify the shocks that affect the economy. They can do this by boosting lending when growth is strong or reducing lending when growth is weak. This generates another type of accelerator effect: the **financial accelerator**.

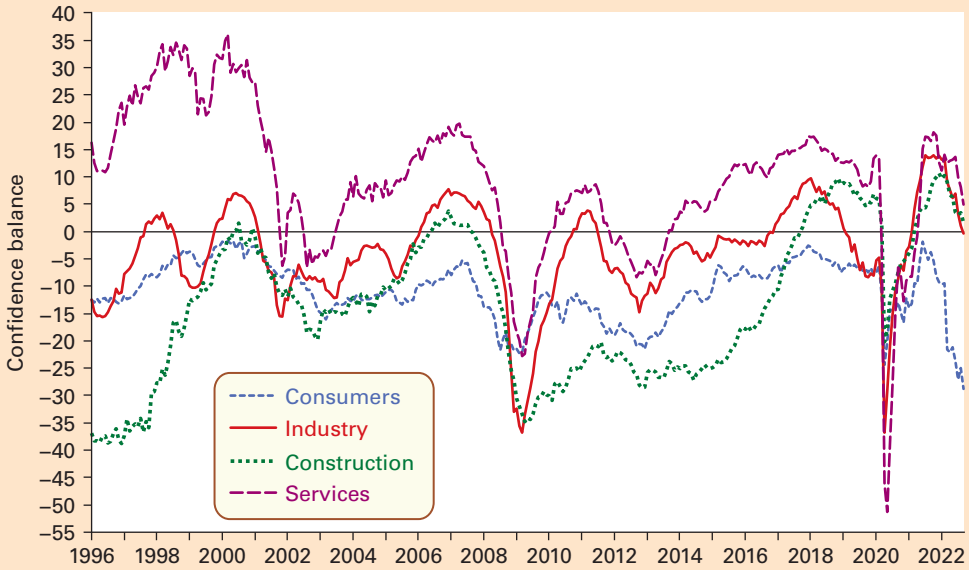
TC 14 p 249

Just as the investment accelerator interacts with the multiplier, so too does the financial accelerator.

BOX 9.3 CONFIDENCE AND SPENDING

Does confidence help to forecast spending?

(a) Confidence indicators in the eurozone



Note: Eurozone figures are the weighted average of the 19 countries (as of 2020) using the euro

Source: Based on data from *Business and Consumer Surveys* (European Commission, DGECFIN)

Economists identify confidence as an important influence on expenditure decisions because it affects the willingness of people and business to spend. Changes in confidence can therefore be a source and an amplifier of economic volatility. But is confidence something tangible that can be measured?

TC 4 p 13

Measures of confidence

Each month, on behalf of the European Commission, consumers and firms across Europe are asked a series of questions, the answers to which are used to compile indicators of consumer and business confidence. For instance, consumers are asked about how they expect their financial position to change. They are offered various options such as 'get a lot better', 'get a lot worse' and balances are then calculated on the basis of positive and negative replies.[1]

[1]More information on EU business and consumer surveys can be found at http://ec.europa.eu/economy_finance/db_indicators/surveys/index_en.htm.

Chart (a) plots the confidence indicators for consumers, industry (manufacturing), construction and services (excluding retail trade) in the eurozone since 1996. The chart captures the volatility of economic sentiment. This volatility is generally greater amongst businesses than consumers. However, confidence fell dramatically across all groups during the financial crisis and the COVID-19 pandemic. It fell again in 2022 as the cost of living soared, with consumers particularly pessimistic.

Confidence and expenditure

Now compare the volatility of confidence in Chart (a) with the annual rates of growth in household consumption and investment in Chart (b). You can see that volatility in economic sentiment is reflected in patterns of both consumer and investment expenditure, although investment is significantly more volatile.

For example, an increase in aggregate demand, whether through a rise in injections (J) or a fall in withdrawals (W), leads to a multiplied rise in national income. But, rising national income may encourage banks to ease credit conditions and so provide more and/or cheaper credit, which itself increases aggregate demand.

In contrast, an economy experiencing a contraction of national income, might see banks tighten credit conditions. This, of course, weakens aggregate demand, further amplifying the contraction.

Pause for thought

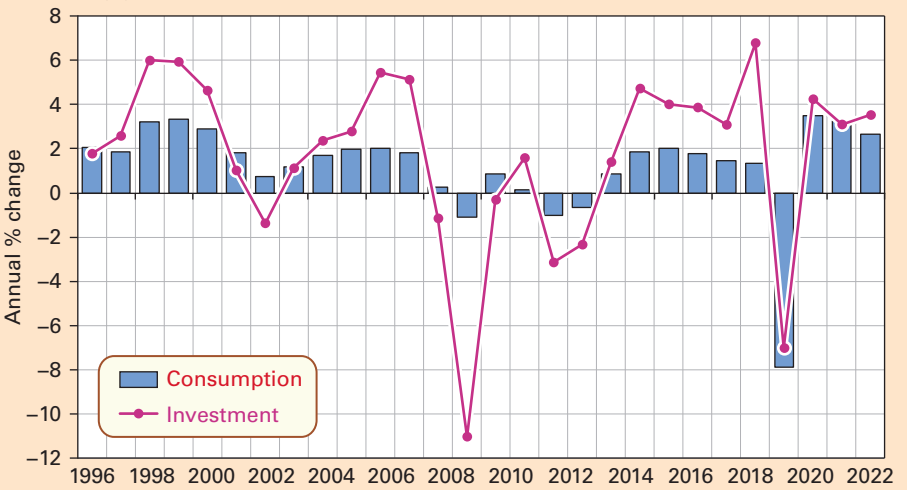
Can the financial and investment accelerators interact?

Why do booms and recessions come to an end? What determines the turning points?

We have examined a variety of reasons why aggregate demand fluctuates. But why, once the economy

CASE STUDIES & APPLICATIONS

(b) Annual change in consumption and investment in the eurozone



Note: Figures from 2022 based on forecasts; eurozone figures are the weighted average of the 19 countries (as of 2022) using the euro

Source: Based on data in *AMECO Database* (European Commission, DGECFIN, May 2022)

TC 11 p 61

What is less clear is the extent to which changes in sentiment *lead* to changes in spending . In fact, a likely scenario is that spending and confidence interact. High rates of spending growth may result in high confidence through economic growth, which in turn leads to more spending. The reverse is the case when economic growth is subdued: low spending growth leads to a lack of confidence, which results in low spending growth and so low rates of economic growth.

TC 14 p 249

Therefore, while confidence may be a source of volatility it may also be part of the process by which shocks are transmitted through the economy. Consequently, it may amplify the peaks and troughs of the business cycle and contribute to the persistence of booms and recessions.

What makes measures of confidence particularly useful is their timeliness. They are released monthly with a delay of only a few days and reflect actions that consumers and investors are likely to take. Although technological progress has facilitated monthly (as opposed to quarterly) measures of GDP and spending, there remains a time lag in their publication.

1. What factors are likely to influence the confidence of (i) consumers; and (ii) businesses?
2. Can consumers become more optimistic while businesses become more pessimistic, and vice versa?

Using time series data from the European Commission based on business and consumer survey data, plot a line chart to show the path of consumer confidence for any two countries of your choice. Describe the patterns you observe noting any similarities or differences between the consumer confidence profile of the two countries.

is booming, does the boom come to an end? Why does a recession not go on for ever? What determines these turning points?

Definitions

Balance sheet recession An economic downturn caused by financially distressed economic agents.

Financial accelerator When a change in national income is amplified by changes in the financial sector, such as changes in interest rate differentials or the willingness of banks to lend.

Ceilings and floors. Actual output can go on growing more rapidly than potential output only as long as there is slack in the economy. As full employment is approached and as more and more firms reach full capacity, so a ceiling to output is reached.

At the other extreme, there is a basic minimum level of consumption that people tend to maintain. During a recession, people may not buy much in the way of luxury and durable goods, but they will still continue to buy food and other basic goods. There is thus a floor to consumption.

Echo effects. Durable consumer goods and capital equipment may last several years, but eventually they will need replacing. The replacement of goods and capital purchased in a previous boom may help to bring a recession to an end.

The accelerator. For investment to continue rising, consumer demand must rise at a faster and faster rate. If this does not happen, investment will fall back and the boom will break.

TC 11 p 61

Sentiment and expectations. Sentiment may change. People may start believing that a boom or recession will not last for ever.

Random shocks. National or international political, social or natural events can affect the mood and attitudes of firms, governments and consumers, and thus affect aggregate demand.

Changes in government policy. In a boom, a government may become most worried by inflation and balance of trade deficits and thus pursue contractionary policies. In a recession, it may become most worried by unemployment and lack of growth and thus pursue expansionary policies. These government policies, if successful, will bring about a turning point in the cycle. This was the hope for the expansionary policies pursued by many governments during the global recession of the late 2000s and in response to the pandemic in 2020–21.

Pause for thought

Why is it difficult to predict precisely when a recession will come to an end and the economy will start growing rapidly?

Recap

1. The consumption cycle mirrors closely the economy's business cycle. Household consumption is thought to depend primarily on current and expected future disposable incomes. The financial system enables households to smooth their consumption by shifting incomes across their lifetimes through borrowing or saving.
2. Household consumption may be affected by a series of 'other determinants' which include: constraints on borrowing, confidence levels, economic uncertainty and a series of balance sheet effects.
3. Investment expenditure is highly volatile. The accelerator theory can help to explain this. It relates the level of investment to *changes* in national income and consumer demand. An initial increase in consumer demand can result in a very large percentage increase in investment; but as soon as the rise in consumer demand begins to level off, investment will fall; and even a slight fall in consumer demand can reduce investment to virtually zero. The accelerator interacts with the multiplier, each amplifying the other.
4. Investment is also affected by a series of 'other determinants', some of which, like access to finance (borrowing constraints) and economic uncertainty also affect household consumption. The scale and cost of investment projects are likely to mean that expectations of the future market environment and business confidence are particularly important determinants.
5. The financial system can help shape or even create the business cycle. A balance sheet recession can arise from increased levels of indebtedness and financial distress. What is more, there may be a financial accelerator effect, whereby financial institutions lend more when growth is strong and less when growth is weak.
6. The interaction of the investment accelerator, the financial accelerator and the multiplier illustrate how the peaks and troughs of the business cycle can be amplified.
7. Booms and recessions do not last for ever. Turning points can be induced by various ceilings and floors, echo effects, the nature of the accelerator, changes in sentiment, random shocks and government policy measures.

9.7 APPENDIX MEASURING NATIONAL INCOME AND OUTPUT

Three routes: one destination

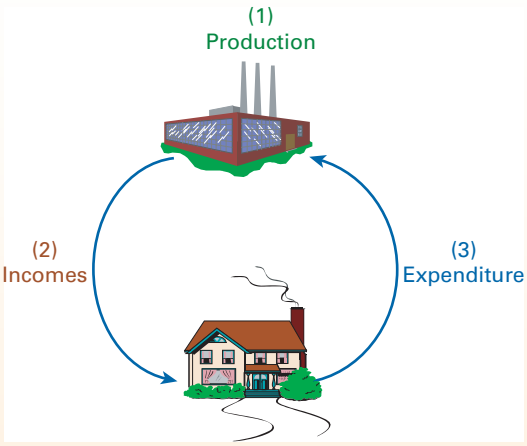
To assess how fast the economy has grown we must have a means of *measuring* the value of the nation's output. The measure we use is *gross domestic product* (GDP).

GDP can be calculated in three different ways, which should all result in the same figure. These three methods are illustrated in the simplified circular flow of income shown in Figure 9.17.

The product method

The first method of measuring GDP is to add up the value of all the goods and services produced in the country, industry by industry. In other words, we focus on firms and add up all their production. Thus method number one is known as the *product method*.

Figure 9.17 The circular flow of national income and expenditure



In the national accounts these figures are grouped together into broad categories such as manufacturing, construction and distribution. The figures for the UK economy for 2020 are shown in the top part of Figure 9.18.

When we add up the output of various firms we must be careful to avoid *double counting*. For example, if a manufacturer sells a television to a retailer for £600 and the retailer sells it to the consumer for £800, how much has this television contributed to GDP? The answer is *not* £1400. We do not add the £600 received by the manufacturer to the £800 received by the retailer: that would be double counting. Instead we either just count the final value (£800) or the value added at each stage (£600 by the manufacturer + £200 by the retailer).

The sum of all the values added by all the various industries in the economy is known as **gross value added at basic prices (GVA)**.

How do we get from GVA to GDP? The answer has to do with taxes and subsidies on products. Taxes paid on goods and services (such as VAT and duties on petrol and alcohol) and any subsidies on products are *excluded* from gross value added (GVA), since they are not part of the value added in production. Nevertheless, GDP is commonly measured at *market prices*: i.e. at the prices actually paid at each stage of production.

GDP at market prices (sometimes referred to simply as GDP) is GVA *plus* taxes on products *minus* subsidies on products. This is illustrated in the bottom part of Figure 9.18.

The income method

The second approach is to focus on the incomes generated from the production of goods and services. This must be the same as the sum of all values added, since value added is simply the difference between a firm's revenue from sales and the costs of its purchases from other firms. This difference is made up of wages and salaries, rent, interest and profit: the incomes earned by those involved in the production process.

Since GVA is the sum of all values added, it must also be the sum of all incomes generated: the sum of wages and salaries, rent, interest and profit.

The second panel in Figure 9.18 shows how these incomes are grouped together in the official statistics. By far the largest category is 'compensation of employees': in other words, wages and salaries. As you can see, the total is the same as in the top part of the figure, although the components are quite different. In other words, GDP is the same whether calculated by the product or the income method.

Note that we do not include transfer payments such as social security benefits and pensions. Since these are not payments for the production of goods and services, they are excluded from GVA. Conversely, part of people's gross income is paid in income taxes. Since it is this gross (pre-tax) income that arises from the production of goods and services, we count wages, profits, interest and rent *before* the deduction of income taxes.

As with the product approach, if we are working out GVA, we measure incomes before the payment of taxes on products or the receipt of subsidies on products, since it is these pre-tax-and-subsidy incomes that arise from the value added by production. When working out GDP, however, we add in these taxes and subtract these subsidies to arrive at a *market price* valuation.

The expenditure method

The final approach to calculating GDP is to add up all expenditure on final output (which will be at market prices). This will include the following:

- Consumer expenditure (*C*). This includes all expenditure on goods and services by households and by

Definitions

Gross value added (GVA) at basic prices The sum of all the values added by all industries in the economy over a year. The figures exclude taxes on products (such as VAT) and include subsidies on products.

Gross domestic product (GDP) (at market prices) The value of output produced within a country over a 12-month period in terms of the prices actually paid.

GDP = GVA + taxes on products − subsidies on products

Figure 9.18 UK GDP: 2020

UK GVA (product based measure): 2020	£m	% of GVA
Agriculture, forestry and fishing	12 578	0.6
Mining and quarrying	13 347	0.7
Manufacturing	186 948	9.6
Electricty, gas, steam and air conditioning supply	30 396	1.6
Water supply, sewerage, waste management and remediation	24 123	1.2
Construction	112 624	5.8
Distribution, transport, hotels and restaurants	310 074	15.9
Information and communications	123 269	6.3
Financial and insurance	168 000	8.6
Real estate activities	266 621	13.7
Professional and support services	237 696	12.2
Total government, health and education	408 294	20.9
Other services and miscellaneous	55 635	2.9
GVA (gross value added at basic prices)	**1 949 605**	**100.0**

UK GVA by category of income: 2020

Compensation of employees (wages and salaries)	1 126 046	57.8
Operating surplus of corporations (gross profit, rent and interest)	489 228	25.1
Mixed income & gross operating surplus of non-corporate sector	391 667	20.1
Tax less subsidies on production (other than those on products)	−57 346	2.9
GVA (gross value added at basic prices)	**1 949 605**	**100.0**

UK GDP: 2020

GVA (gross value added at basic prices)	1 949 605	
plus VAT and other taxes on products	223 969	
less Subsidies on products	−17 501	
GDP (at market prices)	**2 156 073**	

Source: *UK National Accounts, The Blue Book: 2021* (ONS, October 2021)

non-profit institutions serving households (NPISH) (e.g. clubs and societies).

- Government expenditure (G). This includes central and local government expenditure on final goods and services. Note that it includes non-marketed services, such as health and education, but excludes transfer payments, such as pensions and social security payments.
- Investment expenditure (I). This includes investment in capital, such as buildings and machinery. It also includes the value of any increase (+) or decrease (−) in inventories (stocks), whether of raw materials, semi-finished goods or finished goods.
- Exports of goods and services (X).
- Imports of goods and services (M). These have to be *subtracted* from the total in order to leave just the expenditure on *domestic* product. In other words, we subtract the part of consumer expenditure, government expenditure and investment that goes on imports. We also subtract the imported component (e.g. raw materials) from exports.

$$\text{GDP(at market prices)} = C + I + G + X - M$$

Table 9.4 shows the calculation of UK GDP by the expenditure approach.

Table 9.4 UK GDP at market prices by category of expenditure, 2020

	£ million	% of GDP
Consumption expenditure of households and NPISH (*C*)	1 313 295	60.9
Government final consumption (*G*)	480 374	22.3
Gross capital formation (*I*)	360 958	16.7
Exports of goods and services (*X*)	600 973	27.9
Imports of goods and services (*M*)	−596 693	−27.7
Statistical discrepancy	−2834	−0.1
GDP at market prices	**2 156 073**	**100.0**

Source: *UK Economic Accounts time series dataset* (ONS)

Table 9.5 UK GDP, GNY and NNY and households' disposable income at market prices: 2020

	£ million
Gross domestic product (GDP)	**2 156 073**
Plus net income from abroad	−31 996
Gross national income (GNY)	**2 124 078**
Less capital consumption (depreciation)	346 008
Net national income (NNY)	**1 778 069**
Households' disposable income[1]	**1 438 237**

[1]Excluding NPISH sector

Source: *UK Economic Accounts time series dataset* and *Blue Book time series dataset* (ONS)

From GDP to national income

Gross national income

Some of the incomes earned in the country will go abroad. These include wages, interest, profit and rent earned in this country by foreign residents and remitted abroad, and taxes on production paid to foreign governments and institutions (e.g. the EU). On the other hand, some of the incomes earned by domestic residents will come from abroad. Again, these can be in the form of wages, interest, profit or rent, or in the form of subsidies received from governments or institutions abroad.

Gross *domestic* product, however, is concerned with those incomes generated *within* the country, irrespective of ownership. If, then, we are to take 'net income from abroad' into account (i.e. these inflows minus outflows), we need a new measure. This is **gross national income** (**GNY**).[6] It is defined as follows:

GNY at market prices = GDP at market prices + Net income from abroad

Thus GDP focuses on the value of domestic production, whereas GNY focuses on the value of incomes earned by domestic residents.

Net national income

The measures we have used so far ignore the fact that each year some of the country's capital equipment will wear out or become obsolete: in other words, they ignore *capital* **depreciation**. If we subtract an allowance for depreciation (or 'capital consumption') we **get net national income** (**NNY**).

KI 23 p 173

NNY at market prices = GNY at market prices − Depreciation

Table 9.5 shows GDP, GNY and NNY figures for the UK.

Households' disposable income

Households' disposable income measures the income people have available for spending (or saving): i.e. after any deductions for income tax, national insurance, etc. have been made. It is the best measure to use if we want to see how changes in household income affect consumption.

How do we get from GNY at market prices to households' disposable income? We start with the incomes that firms receive[7] from production (plus income from abroad) and then deduct that part of their income that is *not* distributed to households. This means that we must deduct taxes that firms pay – taxes on goods and services (such as VAT), taxes on profits (such as corporation tax) and any other taxes – and add in any subsidies they receive. We must then subtract allowances for depreciation and any undistributed profits. This gives us the gross income that households receive from firms in the form of wages, salaries, rent, interest and distributed profits.

To get from gross income to what is available for households to spend, we must subtract the money that households pay in income taxes and national insurance contributions, but add all benefits to households such as pensions and child benefit.

Households' disposable income is shown at the bottom of Table 9.5.

Definitions

Gross national income (**GNY**) GDP plus net income from abroad.

Depreciation The decline in value of capital equipment due to age or wear and tear.

Net national income (**NNY**) GNY minus depreciation.

[6]In the official statistics, this is referred to as *GNI*. We use *Y* to stand for income, however, to avoid confusion with investment.

[7]We also include income from any public-sector production of goods or services (e.g. health and education) and production by non-profit institutions serving households.

Pause for thought

1. *Should we include the sale of used items in the GDP statistics? For example, if you sell your car to a garage for £2000 and it then sells it to someone else for £2500, has this added £2500 to GDP, or nothing at all, or merely the value that the garage adds to the car: i.e. £500?*
2. *What items are excluded from national income statistics which would be important to take account of if we were to get a true indication of a country's standard of living?*

Households' disposable income =
GNY at market prices − Taxes paid by firms
+ Subsidies received by firms − Depreciation
− Undistributed profits − Personal taxes + Benefits

Taking account of inflation

If we are to make a sensible comparison of one year's national income with another, we must take inflation into account. For example, if this year national income is 10 per cent higher than last year, but at the same time prices are also 10 per cent higher, then the average person will be no better off at all. There has been no *real* increase in income.

As we saw in Section 1.4, *nominal* GDP (or 'money GDP') measures GDP in the prices ruling at the time and thus takes no account of inflation. *Real* GDP, however, measures GDP in the prices that ruled in the *base year*. By comparing real GDP from period to period we eliminate increases in money GDP that were merely due to an increase in prices.

BOX 9.4 **MAKING SENSE OF NOMINAL AND REAL GDP** CASE STUDIES & APPLICATIONS

The interesting case of nominal and real Japanese GDP

TC 15 p 287

When a country is experiencing inflation, nominal GDP will rise faster than real GDP. The reason is that part of the rise in nominal GDP can be explained simply by the rise in prices.

The chart shows nominal and real GDP in Japan and the UK. For both countries the reference year is 2019 and thus in each country nominal and real GDP are the same in that year.

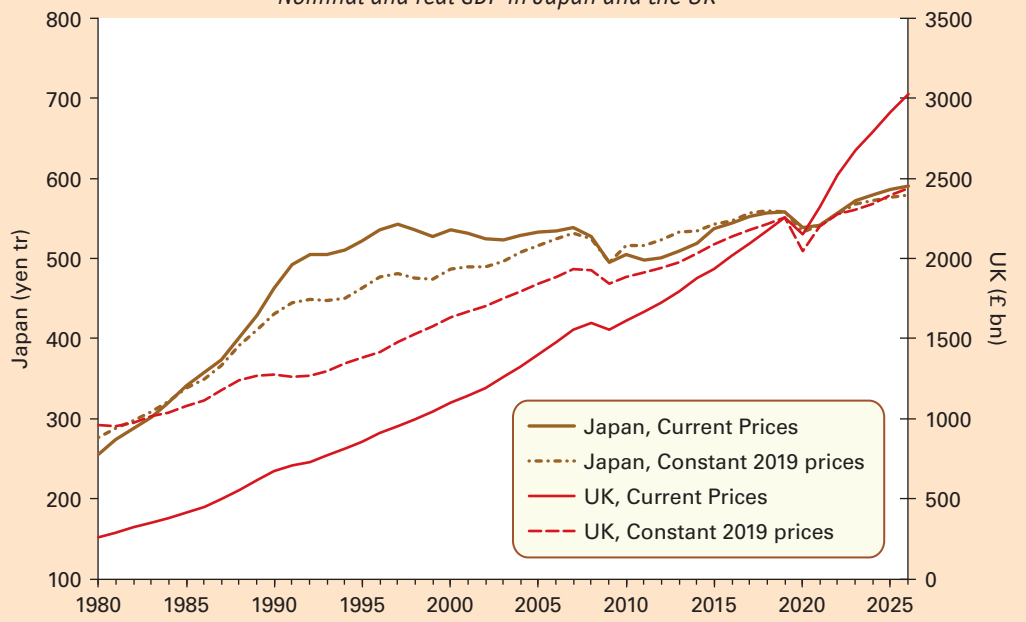
The UK experienced inflation throughout the period. Thus nominal GDP grew faster than real GDP. This can be seen from the graph, where before 2019, nominal GDP is *below* GDP at constant 2019 prices and after 2019, nominal GDP is *above* GDP at constant 2019 prices.

In Japan, however, things have been different with a period of prolonged deflation (negative inflation) since the mid-1990s. The average price of Japanese output typically fell by about 1 per cent each year between 1995 and 2013. Thus, by applying the 2019 reference year to create constant-price values, the effect is to make these numbers typically smaller than the current-price values for much of the period up to the early 2010s.

1. *If a country experiences a consistent rise in the average price level of domestically produced goods would we expect yearly real rates of economic growth to be higher or lower than nominal rates of growth? Explain your answer.*
2. *What effect would re-basing real GDP figures to a later year have on the figures if the country was experiencing (a) inflation; (b) deflation?*

Using time series data from IMF World Economic Outlook Database, construct a line chart similar to that in this Box for any country or country group showing the levels of nominal and real GDP over time.

Nominal and real GDP in Japan and the UK



Note: Data from 2022 are based on forecasts

Source: *World Economic Outlook Database* (IMF, April 2022)

QUESTIONS

1. The table at the bottom of the page shows index numbers for real GDP (national output) for various countries (2015 = 100).

	2015	2016	2017	2018	2019	2020	2021
Australia	100.00	102.70	105.14	108.06	110.21	107.81	112.87
France	100.00	101.03	103.49	105.40	107.34	98.76	105.65
Germany	100.00	102.23	104.97	106.11	107.22	102.33	105.19
Hong Kong	100.00	102.18	106.05	109.07	107.24	100.28	106.71
Italy	100.00	101.29	102.98	103.94	104.46	95.03	101.33
Japan	100.00	100.75	102.44	103.04	102.79	98.17	99.76
Singapore	100.00	103.56	108.39	112.36	113.59	108.88	117.17
UK	100.00	102.26	104.45	106.17	107.95	97.94	105.23
USA	100.00	101.67	103.96	107.00	109.44	105.72	111.72

Source: *World Economic Outlook Database* (IMF, April 2022)

   Using the formula $G = (Y_t - Y_{t-1})/Y_{t-1} \times 100$ (where G is the rate of growth, Y is the index number of output, t is any given year and $t - 1$ is the previous year):
   a. Work out the growth rate for each country for each year from 2016 to 2021.
   b. Plot the figures on a graph. Describe the pattern that emerges.
2. How might we assess the financial well-being of households?
3. What is the difference between a rise in the level of prices and a rise in the rate of inflation?
4. At what point of the business cycle is the country now? What do you predict will happen to growth over the next two years? On what basis do you make your prediction?
5. How regular are the cyclical patterns in real GDP during the course of a business cycle? In what ways could these patterns vary?
6. Identify the four main categories of purchasers of the goods and services produced within a country. Which of these groups of purchasers is the most significant in value terms?
7. For simplicity, taxes are shown as being withdrawn from the inner flow of the circular flow of income (see Figure 9.7 on page 245) at just one point. In practice, different taxes are withdrawn at different points. At what point of the flow would the following be paid: (a) income taxes people pay on the dividends they receive on shares; (b) VAT; (c) business rates; (d) employees' national insurance contributions?
8. In terms of the UK circular flow of income, are the following net injections, net withdrawals or neither? If there is uncertainty, explain your assumptions.
   a. Firms are forced to take a cut in profits in order to give a pay rise.
   b. Firms spend money on research.
   c. The government increases personal tax allowances.
   d. The general public invests more money in building societies.
   e. UK investors earn higher dividends on overseas investments.
   f. The government purchases US military aircraft.
   g. People draw on their savings to finance holidays abroad.
   h. People draw on their savings to finance holidays in the UK.
   i. The government runs a budget deficit (spends more than it receives in tax revenues) and finances it by borrowing from the general public.
   j. The government runs a budget deficit and finances it by printing more money.
   k. As consumer confidence rises, households decrease their precautionary saving.
9. An economy is currently in equilibrium. The following figures refer to elements in its national income accounts.

	£bn
Consumption (total)	1200
Investment	100
Government expenditure	160
Imports	200
Exports	140

   a. What is the current equilibrium level of national income?
   b. What is the level of injections?
   c. What is the level of withdrawals?
   d. Assuming that tax revenues are £140 billion, how much is the level of saving?
   e. If national income now rises to £1600 billion and, as a result, the consumption of domestically produced goods rises to £1160 billion, what is the mpc_d?
   f. What is the value of the multiplier?
   g. Given an initial level of national income of £1600 billion, now assume that spending on exports rises by £80 billion, spending on investment rises by £20 billion and government expenditure falls by £40 billion. By how much will national income change?
10. What is the relationship between the mpc_d and the *mpw*?
11. Assume that the multiplier has a value of 3. Now assume that the government decides to increase aggregate demand in an attempt to reduce unemployment. It raises government expenditure by £100 million with no increase in taxes. Firms, anticipating a rise in their sales, increase investment by £200 million, of which £50 million consists of purchases of foreign machinery. How much will national income rise? (Assume no other changes in injections.)
12. On a Keynesian diagram, draw three *W* lines of different slopes, all crossing the *J* line at the same point. Now draw a second *J* line above the first. Mark the original equilibrium and all the new ones corresponding to each of the *W* lines. Using this diagram, show how the size of the multiplier varies with the *mpw*.
13. Why does the slope of the *E* line in a Keynesian diagram equal the mpc_d? (Clue: draw an mpc_d line.)

14. On a Keynesian diagram, draw two *E* lines of different slopes, both crossing the *Y* line at the same point. Now draw another two *E* lines, parallel with the first two and crossing each other vertically above the point where the first two crossed. Using this diagram, show how the size of the multiplier varies with the mpc_d.
15. What factors could explain why some countries have a higher multiplier than others?
16. What is meant by the term financialisation? Of what significance might this be for the determinants of the business cycle?
17. Explain how the interaction of the multiplier and the investment accelerator might impact on the peaks and troughs of the business cycle.
18. What is the financial accelerator? How might this concept be related to the principle of cumulative causation?
19. In 1974 the UK economy shrank by 2.5 per cent before shrinking by a further 1.5 per cent in 1975. However, the figures for GDP showed a rise of 13 per cent in 1974 and 24 per cent in 1975. What explains these apparently contradictory data?

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

9.1 **John Maynard Keynes (1883–1946).** A profile of the great economist.

9.2 **The Keynesian revolution.** How Keynes' ideas revolutionised the approach to recession and mass unemployment and became economic orthodoxy in the 1950s and 60s.

9.3 **Keynes' views on the consumption function.** An analysis of how the assumptions made by Keynes affect the shape of the consumption function.

9.4 **The paradox of thrift.** How saving more can make the country worse off.

9.5 **The relationship between income and consumption.** Three alternative views of the consumption function.

9.6 **How does consumption behave?** The case looks at evidence on the relationship between consumption and disposable income from the 1950s to the current day.

9.7 **Deriving the multiplier formula.** Using simple algebra to show how the multiplier formula is derived.

9.8 **The explosion of UK household debt.** The growth of household debt in the UK since the mid 1990s and its potential impact on consumption.

9.9 **An international comparison of household wealth and indebtedness.** An examination of households' financial assets and liabilities relative to disposable income in seven developed countries (the G7).

9.10 **Trends in housing equity withdrawal (HEW).** An analysis of the patterns in HEW and consumer spending.

9.11 **Business expectations and their effect on investment.** An examination of business surveys in Europe and the effects of business sentiment on investment.

9.12 **The accelerator.** A numerical analysis of the investment accelerator and the instability of investment.

9.13 **Has there been an accelerator effect since the 1960s?** An examination of the evidence for an accelerator effect in the UK.

9.14 **The multiplier/accelerator interaction.** A numerical example showing how the interaction of the multiplier and accelerator can cause cycles in economic activity.

9.15 **Modelling the financial accelerator.** This case looks at how we can incorporate the financial accelerator effect into the simple Keynesian model of the economy.

9.16 **The GDP deflator.** An examination of how GDP figures are corrected to take inflation into account.

9.17 **Simon Kuznets and the system of national income accounting.** This looks at the work of Simon Kuznets, who devised the system of national income accounting that is used around the world. It describes some of the patterns of economic growth that he identified.

9.18 **Comparing national income statistics.** The importance of taking the purchasing power of local currencies into account.

9.19 **Taking into account the redistributive effects of growth.** This case shows how figures for economic growth can be adjusted to allow for the fact that poor people's income growth would otherwise count for far less than rich people's.

9.20 **The use of ISEW.** An alternative measure to GDP for estimating economic welfare.

WEB APPENDIX

9.1 **Using GDP statistics.** How well do GDP statistics measure a country's standard of living?

Chapter 10

Aggregate supply and economic growth

In this chapter we focus on the factors that influence the aggregate supply of goods and services to an economy. We begin by constructing the aggregate demand and supply (*AD*/*AS*) model. This has an advantage over the circular flow and simple Keynesian models introduced in Chapter 9: it allows us to consider the impact of demand and supply changes not only on the level of output (real GDP), but also on the general level of prices in the economy.

We then turn to look at the causes of economic growth. Why does an economy's output grow over time and what determines its rate? We focus on the determinants of long-term economic growth: growth over many years. Long-term growth is important because it can raise the general standard of living. A key ingredient of long-term growth is technological progress. Therefore, we finish by considering what can be done to foster technological progress and whether or not this involves an active role for governments.

After studying this chapter, you should be able to answer the following questions:

- What can explain the slopes of the aggregate demand (*AD*) and aggregate supply (*AS*) curves? What variables cause the *AD* and *AS* curves to shift?
- What is the effect on the economy of an increase in spending? Will output increase; will prices increase; or will there be some combination of the two?
- What determines the rate of economic growth over the long term?
- What is meant by labour productivity? Why is it important for long-run growth? What are the sources of growth in labour productivity?
- What is meant by capital accumulation, capital deepening, capital shallowing and capital widening?
- Can economies cease to accumulate capital? What would the consequences be?
- How do changes in (a) investment and (b) technological progress affect the long-term growth rate?
- What factors influence the rate of technological progress?
- What can governments do to increase long-term economic growth?

10.1 THE *AD/AS* MODEL

Moving on from the circular flow model of the economy

In the previous chapter, we assumed that a rise in aggregate demand would be reflected purely in terms of an increase in national output (real GDP) – at least up to the full-employment level of national income. We assumed that prices would not change. In practice, a rise in aggregate demand is likely to lead to a rise not only in GDP, but also in prices throughout the economy.

The problem with both the circular flow model and the simple Keynesian model is that they assume that firms' supply decisions respond to demand. But supply decisions are likely to be affected by more than current levels of demand. In particular, they will be affected by prices and costs.

By developing a framework known as the aggregate demand–aggregate supply (*AD/AS*) model we will be able to consider the interaction of the demand and supply sides of the economy. We can then analyse two important issues: how fluctuations in aggregate demand or aggregate supply affect national income *and* prices in the short term and what the key influences on the rate of growth in economic output over the long term are.

The *AD/AS* model is illustrated in Figure 10.1. Note that the aggregate supply curve we are examining here is the *short-run* curve (*SRAS*), as explained below.

As with demand and supply curves for individual goods, we plot quantity on the horizontal axis, except that now it is the *total quantity of national output*, (real) GDP; and we plot price on the vertical axis, except that now it is the *general* price level. Because the general price level relates to the prices of all domestically produced goods and services, it is also known as the **GDP deflator**.

We now examine each curve in turn.

The aggregate demand curve

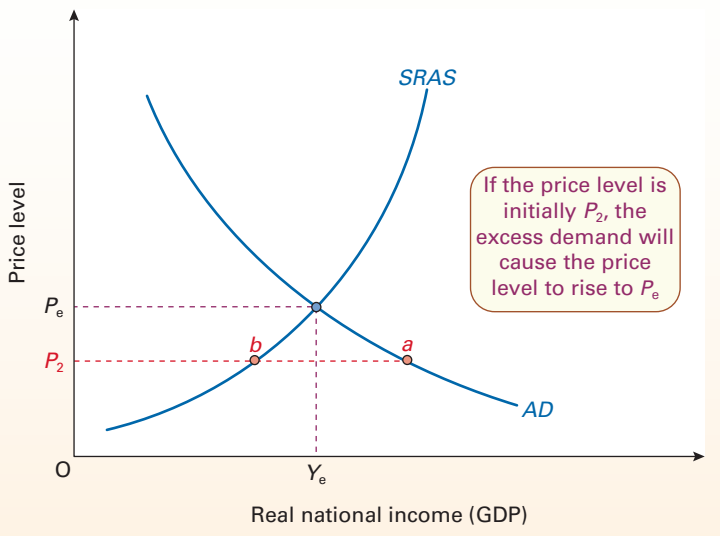
Remember what we said about aggregate demand in Chapter 9. It is the total level of spending on the country's products and consists of four elements: consumer spending on domestic products (C_d), private investment within the country (I_d), government expenditure on domestic products (G_d) and expenditure on the domestically produced element of a country's exports (X_d) (i.e. excluding imported materials and components into those exports). Thus:

$$AD = C_d + I_d + G_d + X_d$$

Definition

GDP deflator The price index of all final domestically produced goods and services (domestic product): i.e. all items that contribute towards GDP.

Figure 10.1 Aggregate demand and supply



The aggregate demand curve shows how much national output (GDP) will be demanded at each level of prices. The level of prices in the economy is shown by a price index (see Web Appendix 9.1 on the student website.).

Slope of the aggregate demand curve

But why will the *AD* curve slope downwards: why will people demand fewer products as prices rise? There are three main reasons.

- **An international substitution effect.** If prices rise, people will be encouraged to buy fewer of the country's products and more imports instead (which are now relatively cheaper); the country will also sell fewer exports (which are now less competitive). Thus imports (a withdrawal) will rise and exports (an injection) will fall. Aggregate demand, therefore, will be lower.
- **An inter-temporal substitution effect.** As prices rise, people will need more money to pay for their purchases. With a given supply of money in the economy, this will have the effect of driving up interest rates (we will explore this in Chapter 11). The effect of higher interest rates will be to discourage borrowing and encourage saving, with individuals postponing current consumption in favour of future consumption. The effect will be a reduction in spending and hence in aggregate demand.
- **Real balance effect.** If prices rise, the value of people's savings will be eroded. They may thus save more (and spend less) to compensate.

TC 6 p 24

The above three effects are *substitution effects* of the rise in prices (see page 31). They involve a switch to *alternatives* – either imports or saving.

TC 5 p 17

There may also be an *income effect*. This will occur when consumers' incomes do not rise as fast as prices, causing a fall in consumers' *real* incomes. Consumers cut down on consumption as they cannot afford to buy so much. Firms, on the other hand, with falling real wage costs, are likely to find their profit per unit rising. However, they are unlikely to spend much more on investment, if at all, as consumer expenditure is falling. The net effect is a fall in aggregate demand.

If, however, consumers' money incomes rise at the *same* rate as prices, there will be no income effect: real incomes have not changed.

Definitions

International substitution effect As prices rise, people at home and abroad buy less of this country's products and more of products from abroad.

Inter-temporal substitution effect Higher prices may lead to higher interest rates and thus less borrowing and more saving.

Real balance effect: As the price level rises, the value of people's money assets falls. They therefore spend less in their attempt to protect the real value of their savings.

Pause for thought

Is the AD curve flatter or steeper the bigger the income and substitution effects?

Shifts in the aggregate demand curve

The *AD* curve can shift inwards (to the right) or outwards (to the left), in exactly the same way as the demand curve for an individual good. A rightward shift represents an increase in aggregate demand, whatever the price level; a leftward shift represents a decrease in aggregate demand, whatever the price level.

A shift in the aggregate demand curve will occur if, for any given price level, there is a change in any of its components – consumption, investment, government expenditure or exports minus imports. Thus if the government decides to spend more, or if consumers spend more as a result of lower taxes, or if business confidence increases so that firms decide to invest more, the *AD* curve will shift to the right.

The aggregate supply curve

The aggregate supply curve shows the amount of goods and services that firms are willing to supply at each level of prices, other things remaining the same. As you can see from Figure 10.1, the short-run aggregate supply (*SRAS*) curve slopes upwards (we explain why below). In other words, the higher the price, the higher the amount that will be supplied.

When drawing the *short-run* aggregate supply curve various things are assumed to remain constant. These include wage rates and other input prices, technology and the total supply of factors of production (land, labour and capital). Because these things obviously do change over time, we have to drop this assumption when drawing *long-run* aggregate supply curves. For now, we concentrate on the short-run *AS* curve.

Why do we assume that wage rates and other input prices are constant in the short run? Wage rates are frequently determined by a process of collective bargaining and, once agreed, will typically be set for a whole year, if not two. Even if they are not determined by collective bargaining, wage rates often change relatively infrequently. So too with the price of other inputs: except in perfect, or near perfect markets (such as the market for various raw materials), firms supplying capital equipment and other inputs tend to change their prices relatively infrequently. They do not

immediately raise them when there is an increase in demand or lower them when demand falls. There is thus a 'stickiness' in both wage rates and the price of many inputs.

Slope of the short-run aggregate supply curve

Why does the short-run aggregate supply curve slope upwards (as in Figure 10.1)? In other words, why will more be produced as the price level rises? The reason is simple: with wages and other input prices being constant, as the prices of firms' products rise, their profitability will rise too. This will encourage firms to produce more.

TC 6 p 24

But what *limits* the increase in aggregate supply in response to an increase in prices? In other words, why is the aggregate supply curve not horizontal? There are two main reasons.

KI 17 p 95

- *Diminishing returns.* With some factors of production fixed in supply, notably capital equipment, firms experience diminishing returns from their other factors, and hence have an upward-sloping marginal cost curve. In microeconomic analysis the upward-sloping cost curves of firms explain why the supply curves of individual goods and services slope upwards. Here in macroeconomics we are adding the supply curves of all goods and services and thus the aggregate supply curve also slopes upwards.
- *Growing shortages of certain variable factors.* As firms collectively produce more, even inputs that can be varied may increasingly become in short supply. Skilled labour may be harder to find, for example.

Thus rising costs explain the upward-sloping short-run aggregate supply curve. The more steeply costs rise as production increases, the less elastic the aggregate supply curve will be. It is likely that, as the level of GDP increases, and as full capacity is approached, so marginal costs will rise faster. The aggregate supply curve will thus tend to get steeper, as shown in Figure 10.1.

Box 10.1 considers further the microeconomic foundations of the *SRAS* curve and their importance to the shape of the curve.

Pause for thought

Is the SRAS *curve flatter or steeper the more significant is the impact of diminishing returns to factors of production?*

Shifts of the aggregate supply curve

The *SRAS* curve will shift if there is a change in any of the variables that are held constant when we plot the curve. Several of these variables, notably technology, the labour force and the stock of capital, change only slowly – normally shifting the curve gradually to the right. This represents an increase in potential output.

Typically, the major cause of leftward/upward shifts in the *SRAS* curve are general rises in input prices, such as wage rates, as these will reduce the amount that firms wish to produce at any level of prices. But supply 'shocks' can have the same effect. For example, lockdowns during the COVID-19 pandemic restricted potential output and thus reduced aggregate supply. Similarly, supply was curtailed by the large rise in energy prices associated with the war in Ukraine.

Equilibrium

The equilibrium price level will be where aggregate demand equals aggregate supply. To demonstrate this, consider what would happen if aggregate demand exceeded short-run aggregate supply: for example, at P_2 in Figure 10.1. The resulting shortages throughout the economy would drive up prices. This would cause a movement up *along* both the *AD* and *SRAS* curves until $AD = SRAS$, at a price level of P_e and a level of national income of Y_e.

TC 8 p 37

Shifts in the AD or SRAS curves

If there is a change in the price level, there will be a movement *along* the *AD* and *SRAS* curves. If any other determinant of *AD* or *SRAS* changes, the respective curve will shift. The analysis here is very similar to shifts and movements along demand and supply curves in individual markets (see pages 33 and 36).

Demand shocks. A positive demand shock means that there is an increase in aggregate demand and the *AD* curve shifts to the right. In contrast a negative demand shock is a decrease in aggregate demand and the *AD* curve shifts to the left.

For example, an increase in consumer confidence (a positive demand shock) leads to a rightward shift of the *AD* curve as shown in Figure 10.2. This will lead to a combination of higher output (Y_1 to Y_2) *and* higher prices (P_1 to P_2). The extent of the increases will depend on the elasticity of the *SRAS* curve: the more elastic the SRAS curve, the more output will rise relative to prices.

KI 8 p 24

The *AD*/*AS* framework shows how, in the face of demand fluctuations, the balance between output and price changes depends on the elasticity of the *SRAS* curve. This is important because a rise in aggregate demand, to the extent that it results in higher prices, will not have a full multiplier effect on real national income. We discuss this further in Section 10.2.

Supply shocks. A positive supply shock means that there is an increase in aggregate supply and the *SRAS* curve shifts downwards to the right.

Figure 10.2 Positive demand shock

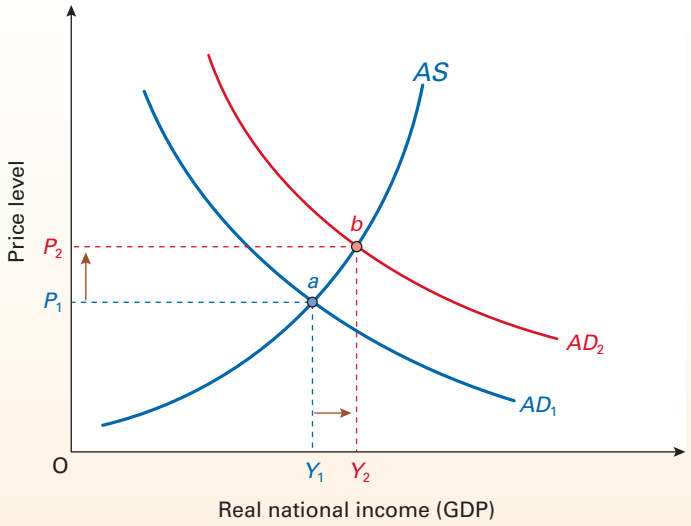
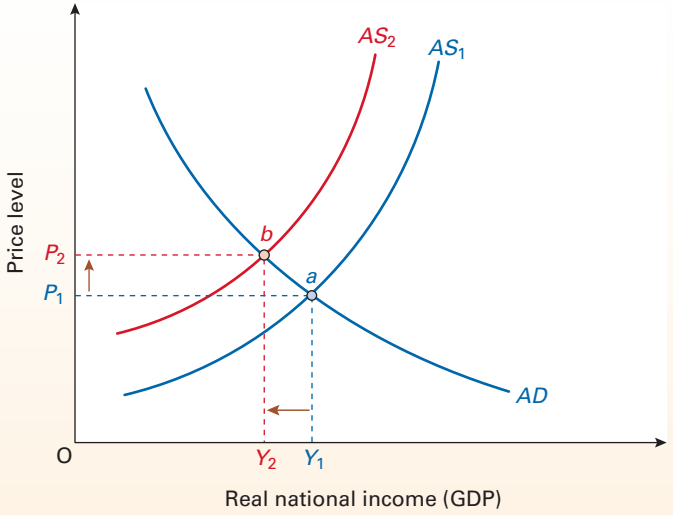


Figure 10.3 Negative supply shock



In contrast a negative supply shock is a decrease in aggregate supply and the *SRAS* curve shifts upwards to the left.

KI 8 p 24

For example, an inflationary shock from rising commodity prices results in a leftward shift of the *SRAS* curve as shown in Figure 10.3. This will lead to a combination of lower output (Y_1 to Y_2) *and* higher prices (P_1 to P_2). The magnitudes will depend on the elasticity of the *AD* curve: the more elastic the *AD* curve, the more output will fall relative to the rise in prices.

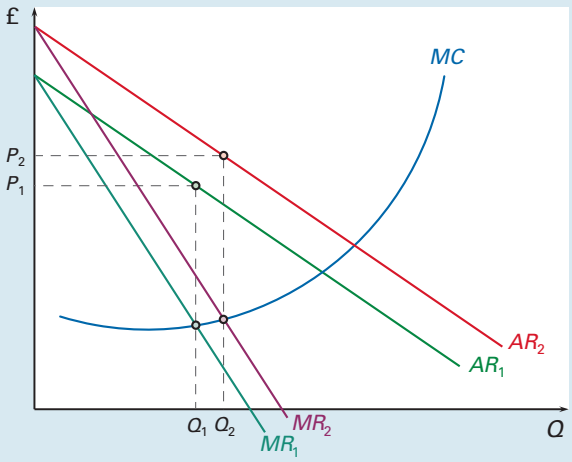
Pause for thought

Why do positive demand shocks increase price levels, but positive supply shocks decrease price levels?

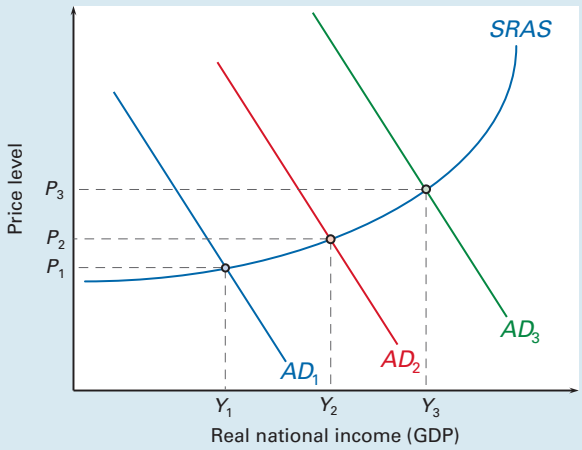
BOX 10.1 SHORT-RUN AGGREGATE SUPPLY

The importance of micro foundations

(a) Short-run response of a profit-maximising firm to a rise in demand



(b) The SRAS curve and the effect of an increase in AD



Recap

1. Equilibrium in the economy occurs where aggregate demand equals aggregate supply.
2. A diagram can be constructed to show aggregate demand and aggregate supply, with the price level on the vertical axis and national output (GDP) on the horizontal axis.
3. The *AD* curve is downward sloping, meaning that aggregate demand will be lower at a higher price level. The reason is that at higher prices there will be substitution effects: (a) there will be more imports and fewer exports; (b) interest rates will tend to be higher, resulting in reduced borrowing and increased saving; (c) people will be encouraged to save more to maintain the value of their savings. Also, if consumer incomes rise less quickly than prices, there will be an income effect too.
4. The *SRAS* curve is upward sloping because the higher prices resulting from higher demand will encourage firms to produce more (assuming that factor prices and technology are fixed).
5. A change in the price level will cause a movement along the *AD* and *SRAS* curves. A change in any other determinant of either *AD* or *SRAS* will cause a shift in the respective curve.
6. While prices and output will tend to move in the same direction following a demand shock, they move in opposite directions following a supply shock.
7. The amount that prices and output change as a result of a shift in aggregate demand will depend on the shape of the *SRAS* curve.

10.2 ALTERNATIVE PERSPECTIVES ON AGGREGATE SUPPLY

How does aggregate supply respond to changes in aggregate demand?

The response of aggregate supply to a change in aggregate demand is keenly debated among economists. The debate can best be understood in terms of the nature of the aggregate supply (*AS*) curve. We will start with the short-run aggregate supply curve and then look at the long-run curve.

The short-run aggregate supply curve

Assume that there is a rise in aggregate demand. The short-run effect on output and prices will depend on the shape of the *SRAS* curve. Let us examine the different analyses of the *SRAS* curve. The various

EXPLORING ECONOMICS

To understand the shape of the short-run *AS* curve, it is necessary to look at its microeconomic foundations. How will *individual* firms and industries respond to a rise in demand? What shape will their individual supply curves be?

TC 10 p 49

In the short run, we assume that firms respond to the rise in demand for their product *without* considering the effects of a general rise in demand on their suppliers or on the economy as a whole. We also assume that the prices of inputs, including wage rates, are constant.

In the case of a profit-maximising firm under monopoly or monopolistic competition, there will be a rise in price and a rise in output. In diagram (a), profit-maximising output rises from where $MC = MR_1$ to where $MC = MR_2$. Just how much price changes compared with output depends on the slope of the marginal cost (*MC*) curve.

The nearer the firm is to full capacity, the steeper the *MC* curve is likely to be. Here the firm is likely to find diminishing returns setting in rapidly, and it is also likely to have to use more overtime with correspondingly higher unit labour costs. If, however, the firm is operating well below capacity, it can probably supply more with little or no increase in price. Its *MC* curve may thus be horizontal at lower levels of output.

KI 17 p 95

When there is a general rise in demand in the economy, the *aggregate* supply response in the short run can be seen as simply the sum of the responses of all the individual firms. The short-run *AS* curve will look something like that in diagram (b).

If there is generally plenty of spare capacity, a rise in aggregate demand (e.g. from AD_1 to AD_2) will have a big effect on output and only a small effect on prices. However, as more and more firms find their costs rising as they get nearer to full capacity, so the *SRAS* curve becomes steeper. Further increases in aggregate demand (e.g. from AD_2 to AD_3) will have bigger effects on prices and smaller effects on output (GDP).

KI 17 p 95

A general rise in prices, of course, means that individual firms were mistaken in assuming that a rise in price from P_1 to P_2 in diagram (b) was a *real* price rise (i.e. relative to prices elsewhere).

TC 3 p 10

1. *What happens to real wage rates as we move up the SRAS curve?*
2. *If markets were to become dominated by oligopolies what effect might this have on the SRAS curve?*

Write a short technical report contrasting the orthodox, extreme Keynesian and new classical views of the SRAS curve.

approaches to analysing aggregate supply are illustrated in Figure 10.4.

The mainstream position

The mainstream or moderate position, also known as the 'orthodox' position, (Figure 10.4(a)) is the one that we have just described in Section 10.1. This depicts the *SRAS* curve as upward sloping. It is assumed that wages and many other input prices exhibit *some* 'stickiness' in the short term. A rise in demand will not simply be absorbed in higher input prices: in other words, output will rise too.

Nevertheless, as more variable factors are used, firms will experience diminishing returns. Marginal costs will rise. The less the spare capacity in firms, the more rapidly marginal costs will rise for any given increase in output and hence the steeper the *SRAS* curve will be.

KI 17 p 95

Therefore, the mainstream view is that an increase in *AD* will have some effect on prices and some effect on output and employment (see Figure 10.2 and 10.4(a). The extent of these effects will depend on the economy's current level of output relative to its potential (normal capacity) output. The higher actual output is relative to potential output, the steeper the *SRAS* becomes. Therefore, aggregate supply will become less and less responsive to aggregate demand as cyclical unemployment continues to fall.

TC 12 p 195

The extreme Keynesian position

The extreme Keynesian position (Figure 10.4(b)) is that up to the level of income that will generate full employment (Y_F) (similar to the concept of potential national income) the *SRAS* curve is *horizontal*. Hence, this is consistent with the simple Keynesian model we developed in Chapter 9. A rise in aggregate demand from AD_1 to AD_2 will raise output from Y_1 to Y_2, but there will be *no effect on prices* until full employment is reached. Thereafter, further increases in aggregate demand are inflationary since output cannot exceed the full-employment level, other than in the very short run.

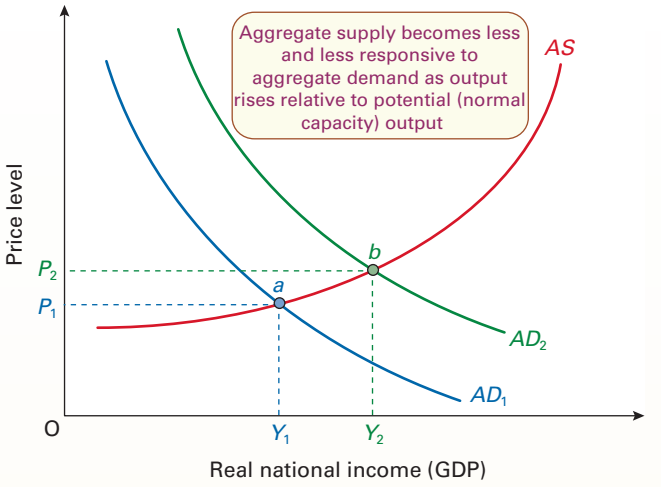
Pause for thought

If the full-employment output level is a 'ceiling' for output, how could output levels exceed it?

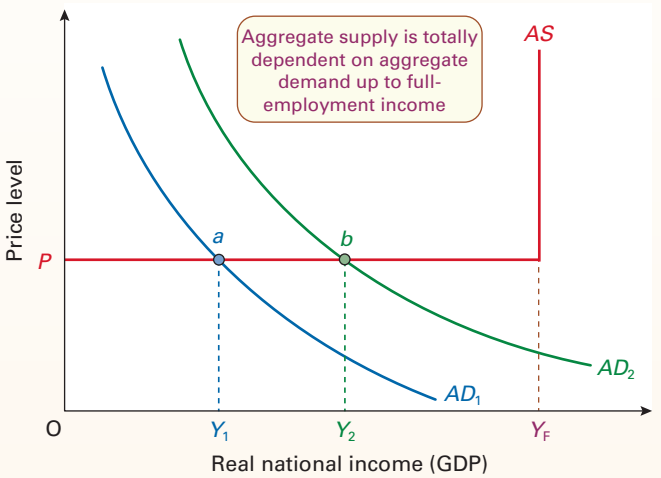
The important point here is that aggregate supply up to the full-employment level is determined entirely by the level of aggregate demand. However, there is no guarantee that aggregate demand will intersect aggregate supply at full employment. Therefore, governments should manage aggregate demand by appropriate fiscal and monetary policies to ensure production at Y_F. We look at various macroeconomic policies in Chapter 13.

TC 7 p 26

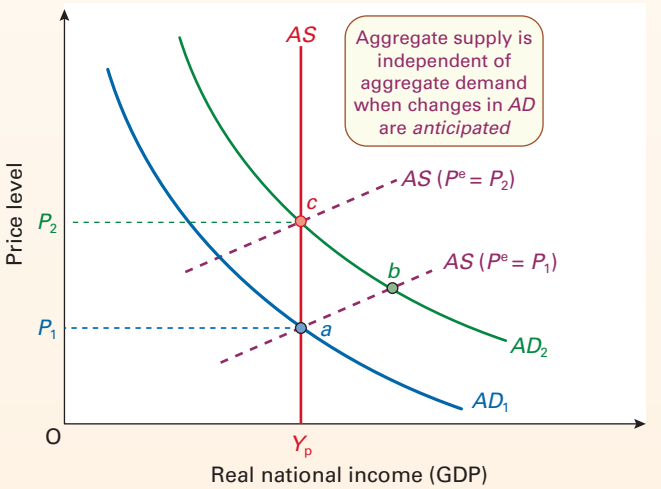
Figure 10.4 Different short-run aggregate supply curves



(a) Mainstream position



(b) Extreme Keynesian position



(c) New classical position

The new classical position

In contrast, economists known as 'new classicists'[1] argue that the *SRAS* curve is *vertical* at potential output (Y_P), as in Figure 10.4(c). This rests on two important assumptions. First, is the assumption of **continuous market clearing**. This means that all markets continuously adjust to their equilibrium. Second, is the assumption of **rational expectations**. This means that people use all available information and predict inflation, or any other macroeconomic variable, as well as they can. The important point here is that forecasting errors are random so that, on average, people's expectations of inflation are correct.

TC 11 p 61

Therefore, anticipated rises in aggregate demand will quickly work through both goods and factor markets into higher prices, not into higher output. There has been no increase in *real* aggregate demand. Output remains at its potential (normal capacity) level Y_P. Thus it is essential to keep (nominal) demand under control if *prices* are to be kept under control.

Unanticipated change in aggregate demand. An upward sloping *SRAS* curve would be observable only if changes in aggregate demand were *unanticipated* and even then deviations in output from its potential level would be transitory. If, in Figure 10.4(c), aggregate demand were to rise unexpectedly, say from AD_1 to AD_2, people would not foresee the upward effect on general prices. Instead, the expected price level (P^e) is P_1($P^e = P_1$). As the general price level rises, firms (wrongly) believe that it is profitable for them to expand output levels, not realising that this will simply drive up wages and other input prices. This is equivalent to the move from *a* to *b* in Figure 10.4(c).

Once people recognise these errors, however, and input prices do rise, output adjusts back to its potential level Y_P. The economy moves from point *b* to *c*. In the presence of rational expectations and continuous market clearing this adjustment is likely to happen relatively quickly.

TC 11 p 61

If the government wants to expand aggregate supply and get more rapid economic growth, it is no good, they argue, concentrating on demand. Instead, governments should concentrate directly on supply by encouraging enterprise and competition, and generally by encouraging markets to operate more freely. For this reason, this approach is often labelled **supply-side economics**. If successful, these will shift the vertical *AS* curve to the right.

[1] They are known as 'new classicists' because their analysis is similar to the pre-Keynesian or 'classical' theory that output is determined by the economy's capacity to produce, not by the level of demand.

Pause for thought

If there were an unexpected decrease in aggregate demand, would new classicists expect output to fall below its potential level? Explain.

The long-run aggregate supply curve

A vertical long-run AS curve

While some new classical economists argue that the *short-run AS* curve is vertical, most economists argue that it is only the *long-run AS* curve that is vertical at the potential level of output (Y_P): see Figure 10.5(b). Any rise in nominal aggregate demand would lead simply to a rise in prices and no long-term increase in output at all.

The mainstream view is that long-term increases in output could occur only through rightward shifts in this vertical long-run *AS* curve, in other words through increases in potential output. To achieve this governments therefore should focus largely on supply-side policy, such as policies which help foster technological progress (see Section 13.4).

But why do these economists argue that the long-run *AS* curve is vertical? They justify this by focusing on the *interdependence of markets*. Assume initially that the economy is operating at the potential level of output (Y_P). Now assume that there is an increase in aggregate demand. This will initially lead firms to raise both prices and output prices in accordance with the short-run *AS* curve. There is a movement from point *a* to point *b* in Figure 10.6. Output rises to Y_2.

However, as raw material and intermediate goods producers raise their prices, so this will raise the costs of production of firms using these inputs. A rise in the price of steel will raise the costs of producing cars and washing machines. At the same time, workers, seeing the prices of goods rising, will demand higher wages. Firms will be relatively willing to grant these wage demands since they are experiencing buoyant demand.

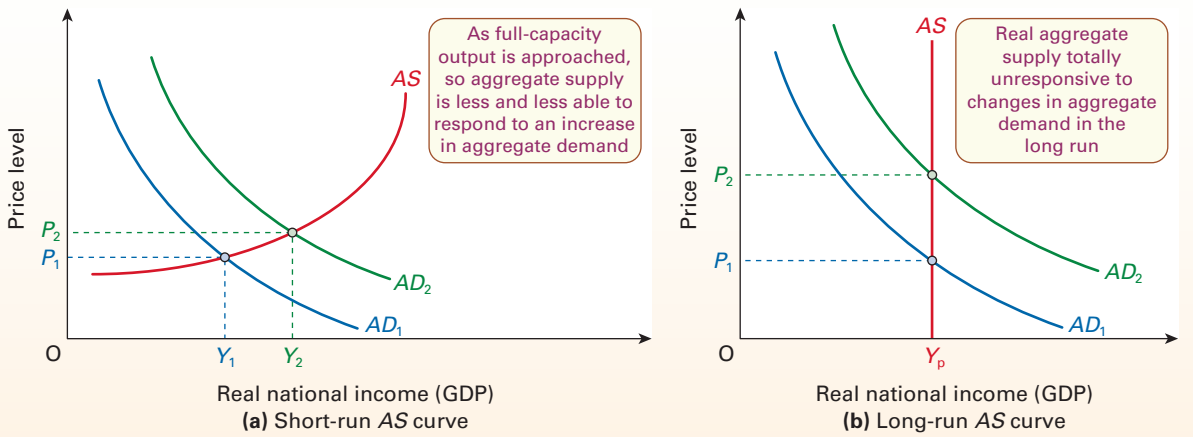
Definitions

Continuous market clearing The assumption that all markets in the economy continuously clear so that the economy is permanently in equilibrium.

Rational expectations Expectations based on the *current* situation. These expectations are based on the information people have to hand. While this information may be imperfect and therefore people will make errors, these errors will be random.

Supply-side economics An approach that focuses directly on aggregate supply and how to shift the aggregate supply curve outwards.

Figure 10.5 Short-run and long-run aggregate supply curves



The effect of all this is to raise firms' *costs*, and hence their prices. As prices rise for any given level of output, so the short-run *AS* curve will shift upward. This is shown by a move to $AS_{2(\text{short run})}$ in Figure 10.6. The economy moves from point *b* to point *c*. Thus output can only temporarily rise above the potential level (Y_P). Although the process may take some time, the more flexible markets are, the more quickly higher costs are likely to be passed through into higher prices.

The long-run effect, therefore, of a rise in aggregate demand from AD_1 to AD_2 is a movement from point *a* to point *c*. The long-run aggregate supply curve passes through these two points. It is vertical at the potential level of output. A rise in aggregate demand will therefore have no long-run effect on output. The entire effect will be felt in terms of higher prices.

An upward-sloping long-run AS curve

Some Keynesian economists, however, argue that the long-run *AS* curve is upward sloping, not vertical. Indeed, it may be even shallower than the short-run curve. For them, potential output is affected by changes in aggregate demand.

The argument here is that with a rise in demand, firms may be encouraged to invest in new plant and machinery. This increase in the stock of capital will increase the *capacity* of the economy to produce. In other words, the increase in aggregate demand will increase potential output. The result is that firms may well be able to increase output significantly in the long run with little or no increase in their prices. Their long-run *MC* curves are much flatter than their short-run *MC* curves.

Figure 10.6 The long-run aggregate supply curve when firms are interdependent

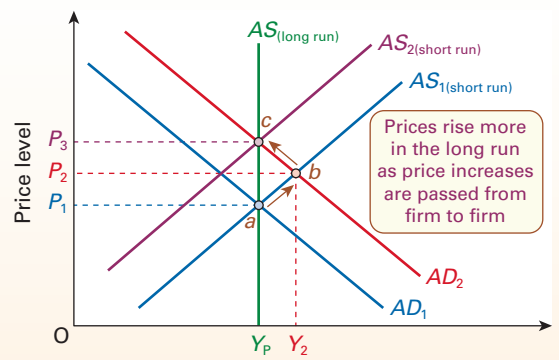
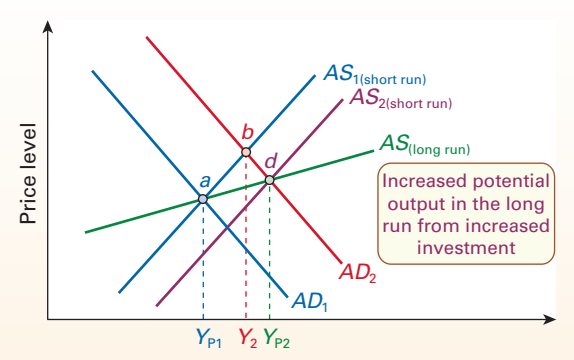


Figure 10.7 Effects of investment on the long-run aggregate supply curve



Again, assume initially that output is at the potential level. In Figure 10.7 this is shown as Y_{P1}. Aggregate demand then increases to AD_2. Equilibrium moves to point *b* with GDP at Y_2. The resulting increased investment shifts the short-run *AS* curve to the right. Equilibrium moves from point *b* to *d*. Point *d* is now at the new potential level of output, Y_{P2}. The long-run *AS* curve thus joins points *a* and *d*.

The way the diagram is drawn, the long-run *AS* curve is more elastic than the short-run curve. There is a relatively large increase in output and a relatively small increase in price. If the rise in costs had been more substantial, curve $AS_{2(\text{short run})}$ could be above curve $AS_{1(\text{short run})}$. In this case, although the long-run *AS* curve would still be upward sloping, it would be steeper than the short-run curves: point *d* would be above point *b*.

Pause for thought

If a shift in the aggregate demand curve from AD_1 to AD_2 in Figure 10.7 causes a movement from point a *to point* d *in the long run, will a shift in the aggregate demand curve from AD_2 to AD_1 cause a movement from point* d *back to point* a *in the long run?*

The long-run *AS* curve will be steeper if the extra investment causes significant shortages of materials, machinery or labour. This is more likely when the economy is already operating near its full-capacity output. It will be flatter, and possibly even downward sloping, if the investment involves the introduction of new cost-reducing technology.

Areas of general agreement

Despite differences between economists on the nature of aggregate supply, there are two general points of agreement that have emerged, at least among the majority of economists.

- In the short run, changes in aggregate demand can have a significant effect on output and employment. If there is a collapse in demand, as happened following the global financial crisis and the COVID-19 pandemic, governments and/or central banks should intervene through expansionary fiscal and/or monetary policies. Only a few extreme new classical economists would disagree with this proposition.
- In the long run, changes in aggregate demand will have much less effect on output and employment and much more effect on prices. In fact, many economists say that there will be no effect at all on output and employment, and that the whole effect will be on prices. There is still a substantial body of Keynesians, however, especially 'post-Keynesians', who argue that changes in aggregate demand will have substantial effects on long-term output and employment via changes in investment and hence in potential output.

The mainstream view generally recognises that, if the economy is in deep recession, it may be necessary to expand aggregate demand. However, macroeconomic policy should not focus exclusively on the demand side. Long-term growth, it is argued, depends primarily on changes in supply (i.e. in potential output). It is important, therefore, for governments to develop an effective supply-side policy if they want to achieve faster long-term economic growth.

TC 7 p 26

The rest of the chapter focuses on long-term growth. It analyses the determinants of economic growth, not over short periods of time, but over many years, perhaps decades or generations.

Recap

1. The impact of changes in aggregate demand on prices (and output) is affected by the nature of aggregate supply (*AS*) curve.
2. The mainstream view is that if nominal aggregate demand changes, then in the short run it is likely to affect real GDP (*Y*) according to the degree of slack in the economy. This is because both wage rates and prices tend to be relatively sticky. The *SRAS* curve is therefore upward sloping but becomes progressively steeper at higher levels of real national income.
3. In the extreme Keynesian model, increases in nominal aggregate demand have a full multiplied effect on real GDP up to the full-employment output level because prices are constant. Thereafter, further increases in aggregate demand are inflationary and increases in real GDP are only possible in the very short term. The *SRAS* curve is therefore a flat horizontal line.
4. Some new classicists argue that the *SRAS* curve may be vertical. This is because of the flexibility of markets and the ability of rational people to forecast the effect of expected changes in aggregate demand on prices.
5. The mainstream view is that the long-run aggregate supply (*LRAS*) curve is vertical because price increases from any rise in aggregate demand tend to be passed on from one firm to another and feed into wage increases.
6. Some argue, however, that the *LRAS* curve may be upward sloping. If a sustained increase in demand leads to increased investment, this can have the effect of shifting the short-run *AS* to the right and making the long-run *AS* curve upward sloping, not vertical.

10.3 INTRODUCTION TO LONG-TERM ECONOMIC GROWTH

What is the historical pattern of economic growth?

Our focus so far has been on economic volatility. This is not surprising given that economies are inherently volatile and often experience significant fluctuations in economic activity. Yet, when we step back and look at the longer span of history, these short-term fluctuations take on less significance. What we see is that economies tend to experience long-term economic growth.

TC 7 p 26

These twin characteristics of growth are nicely captured in Figure 10.8, which plots for the UK both the *level* of real GDP and annual percentage *changes* in real GDP. It shows that while the rate of economic growth is volatile, the volume of output grows over time.

The rate of long-term economic growth in developed nations, such as the UK, has meant that average living standards have improved markedly. When measured in terms of real GDP per head, all developed nations are considerably richer today than they were, say, 50 or 60 years ago.

TC 7 p 26

The picture, however, is not one of universal improvement. People are not necessarily happier; there are many stresses in modern living; the environment is in many respects more polluted; inequality has increased in most countries; for many people, work is more demanding and the working day is longer than in the past; there is more crime and more insecurity. Hence, 'more' is not always 'better'.

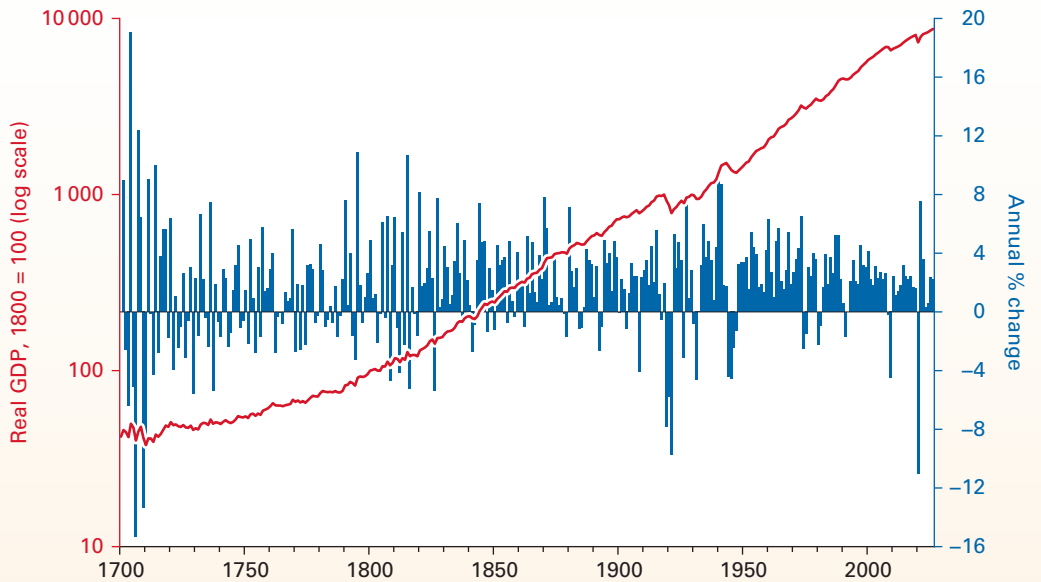
Nevertheless, most people *want* more consumer goods; they want higher incomes. In this chapter, we examine what causes long-term economic growth, and how it can be increased. We leave you to judge whether a materially richer society is a better society.

Long-run growth and the *AD/AS* model

Sustained economic growth over the longer term requires a sustained increase in *potential output*. This means that there has to be a continuous rightward shift in aggregate supply. When viewed through the *AD/AS* model, long-run growth therefore means a continuous rightward shift of the *AS* curve. The rate of growth in potential output is then the rate at which the *AS* curve shifts rightwards.

We can use Figure 10.9 to illustrate in general terms the process of long-run growth. Assume initially that the economy is at equilibrium with output (real national income) at the potential level (Y_{P_1}) and the economy's price level at P_1. Now assume that potential

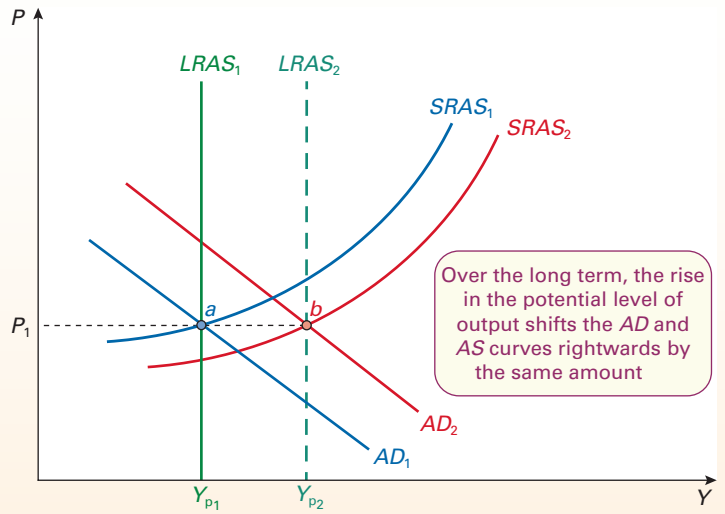
Figure 10.8 Output and economic growth in the UK since 1700



Note: Growth is the annual growth in constant-price GDP

Sources: 1700–1948 based on data from *A Millennium of Macroeconomic Data*, (Bank of England) available at https://www.bankofengland.co.uk/statistics/research-datasets; from 1949 to 2021 based on series ABMI and IHYP (National Statistics); from 2022 based on *IMF, World Economic Outlook* (October 2022)

Figure 10.9 Long-term growth in integrated *AD/AS* framework



output rises from Y_{P_1} to Y_{P_2}. We will look at how this might happen later in the chapter.

The short- and long-run *AS* curves move rightwards – say to $SRAS_2$ and $LRAS_2$. The size of the shifts reflects the increase in the economy's productive capabilities. If the economy's extra capacity is used, it has the effect of increasing the real national income of the country; people are now able to make more purchases. There is also likely to be a rightward shift in the *AD* curve. If the extra real income generated by the increase in potential output generates an equivalent amount of additional real spending, the *AD* curve will shift to AD_2. In such a case, actual output would increase by the same amount as potential output: i.e. $Y_{P_2} - Y_{P_1}$.

If, however, the rise in potential output did not translate into sufficient extra spending, the *AD* curve would not shift sufficiently to the right to give a new equilibrium at Y_{P_2} – a negative output gap would emerge. In such a case, discretionary rises in government spending and/or cuts in taxation (fiscal policy) or cuts in interest rates or increases in the money supply (monetary policy) may be required to shift the *AD* curve to AD_2. We examine fiscal and monetary policies in Chapter 13.

Generally, however, over a time span of several years, periods of deficient demand are likely to be matched by periods of excess demand. In the long run, therefore, rightward shifts in the *AS* curves would be matched by equivalent rightward shifts in the *AD* curve. In other words, in the long run, economies tend towards equilibrium at the potential level of national income as markets and expectations adjust.

Comparing the growth performance of different countries

Long-term growth cannot be explained by a closing of the gap between actual and potential output: by an expansion of aggregate demand leading to a fuller use of resources. Instead, the explanation lies on the supply side. Countries' economic *capacity* has increased.

Table 10.1 shows a series of average economic growth rates: the average annual growth in real GDP (total output), real GDP per worker (a measure of the productivity of labour) and real GDP per capita (real GDP per head of the population). The average growth rates are for several developed countries since the 1960s.

As you can see from Table 10.1, there are considerable differences in the rates of growth experienced by the different countries. The effect of even very small differences can have a significant effect when looked at over many years. This is particularly important when we are thinking about the impact of growth on a country's *living standards*.

Pause for thought

Is it possible to have economic growth without an increase in output per worker? Explain.

Table 10.1 Average annual growth rates (%) 1961–2023

	Real GDP	Real GDP per worker	Real GDP per capita
Ireland	5.0	3.7	4.0
Japan	3.5	2.9	3.0
Portugal	3.1	2.9	2.9
Spain	3.2	2.4	2.4
Norway	3.0	2.0	2.3
Austria	2.6	2.1	2.2
Belgium	2.5	2.0	2.1
France	2.7	2.1	2.0
Italy	2.3	2.0	2.0
Netherlands	2.7	1.4	2.0
USA	3.0	1.6	2.0
Germany	2.3	1.7	2.0
Sweden	2.5	1.9	2.0
UK	2.3	1.7	1.9
Switzerland	2.2	1.1	1.3

Note: Figures from 2022 based on forecasts; German figures based on West Germany only up to 1991

Source: Based on data from *AMECO Database* (European Commission, DGECFIN)

In the context of living standards, it is the growth in real GDP *per capita* that is important (as shown in the last column of the table). An increase in the population will, other things being equal, lower living standards because more people will be sharing a given amount of real national income. Hence, when analysing economic growth over time it is real GDP per capita that we tend to focus on.

Pause for thought

Why may the rates of growth in average output per worker and in average output per head of the population differ?

In general, GDP per capita in the richer developed countries has grown at a slower rate than in the less rich ones. The result has been a narrowing of the gap in living standards. For example, in 1950, GDP per head in the USA (in purchasing-power standard terms) was 2.5 times that in West Germany and 20 times that in Japan. It was estimated that, by 2026, GDP per head in the USA would be only 17 per cent higher than that in Germany and 36 per cent higher than that in Japan.

This **convergence in GDP per head**, however, has not been universal across the world. Although countries such as Brazil, China, India and many other Asian countries have grown very rapidly, and in recent years some of the poorer African countries too, there remain others, often blighted by war or corruption or rapid population growth rates, where real GDP per capita has grown at pitifully slow rates, and in some cases has even declined.

Although recent generations have come to expect economic growth, it is a relatively new phenomenon. For most of the last 2000 years, countries have experienced virtually static output per head over the long term. Economic growth has become significant only once countries have undergone an industrial revolution, and it is only with the technological advances of the twentieth and now the twenty-first centuries that long-term growth rates of 2 per cent or more have been achieved.

The causes of economic growth

If we look back at Table 10.1 we can see the importance of the growth in real GDP (output) per worker for the growth in real GDP per capita and hence for a country's living standards. Over the long term, the rate of growth of the workforce generally reflects population growth – although, with an ageing population, it tends to lag behind somewhat. Therefore, the growth in output per capita is principally the result of increases in output per worker.

Output per worker is a measure of **labour productivity**: it is an indicator of how effective workers are in the production process. Since the growth in output per worker is the key to understanding the growth in output per capita, it then follows that the growth of labour productivity is crucial in determining a country's long-run economic growth.

But, what explains the growth in labour productivity? In fact, there are three key sources of growth in labour productivity:

TC 15 p 287

- An increase in the *quantity of physical capital (K)*. Here we are referring to the accumulation of capital, such as machinery/equipment and office/factory space, that arises through investment. The better equipped workers are, the more productive they will be.

KI 23 p 173

Definitions

Convergence in GDP per head The tendency for less rich developed countries to catch up with richer ones. Convergence does not apply to many of the poorer developing countries, however; the gap between them and richer countries has tended to widen.

Labour productivity Output per unit of labour: for example, output per worker or output per hour worked.

- An increase in **human capital**. Here we are referring to the knowledge, skills, competencies and other attributes of individuals that impact on their ability to produce goods and services and to generate ideas.
- **Technological progress.** Developments of computer technology, of new techniques in engineering, of lighter, stronger and cheaper materials, of digital technology in communications and of more efficient motors have all contributed to a massive increase in the productivity of capital. Machines today can produce much more output than machines in the past that cost the same to manufacture. Therefore workers are able to produce considerably more today, even allowing for the increases that have taken place in physical and human capital.

Box 10.2 discusses the concept of labour productivity further and compares its growth in the UK with that of other developed countries.

In the next two sections, we will examine these sources of growth. But before we go on, a word of caution: long-term economic growth does not mean universal improvement. People are not necessarily happier; there are many stresses in modern living; the environment is in many respects more polluted; inequality has increased in most countries, especially over the past 20 years; for many people work is more demanding, and the working day is longer than in the past; there is more crime and more insecurity. More is not always better.

KEY IDEA 34 TC 15

Long-term growth in a country's output depends on a growth in the quantity and/or productivity of its resources. Potential economic growth depends on a country's resources, technology and productivity. This is crucial to understanding what underlies the wealth of nations and why some countries have faster growth rates than others. It forms the fifteenth of our Threshold Concepts.

Definition

Human capital The qualifications, skills, expertise and well-being that contribute to a worker's productivity.

Recap

1. Economies are inherently volatile in the short run, but most countries experience growth over the long term. The determinants of long-term economic growth lie primarily on the supply side. Long-term growth is therefore consistent with continuous rightwards shifts of the *AS* curve.
2. Most developed countries have experienced average annual rates of economic growth of more than 2 per cent over the past 50 years, but there have been considerable differences between countries.
3. The income gap between developed countries has tended to narrow as the less rich ones have grown faster than the richer ones. Some of the poorest countries of the world, however, have experienced very low rates of growth, with the result that the gap between them and richer countries has widened.
4. When analysing the long-term rate of economic growth, we focus on the growth in real national income per capita (per head of the population). This is because it provides a better indicator of patterns in living standards than does the growth in total real national income.
5. A key determinant of the growth in real national income per capita is the growth in real national income per worker (labour productivity). There are three principal determinants of labour productivity: the quantity of physical capital, the quantity of human capital and technological progress.

10.4 ECONOMIC GROWTH WITHOUT TECHNOLOGICAL PROGRESS

How important is physical capital to economic growth?

Capital accumulation and capital deepening

The ratio of an economy's capital stock to the size of its labour force (K/L) is a measure of its **capital intensity**. As countries accumulate capital they have more manufactured equipment to help in production. This is referred to as **capital accumulation**. When the rate of capital accumulation is greater than the growth of the workforce, then each worker has more capital to work with. This increase in capital intensity is

Definitions

Capital intensity The amount of physical capital that workers have to operate with and which can be measured by the amount of capital per worker (K/L).

Capital accumulation An increase in the amount of capital that an economy has for production.

Capital deepening (shallowing) An increase (a decrease) in the amount of capital per worker (K/L).

BOX 10.2 MEASURING LABOUR PRODUCTIVITY

Mind the UK productivity gap

TC 15 p 287

Long-term increases in real GDP per capita depend crucially on the growth in labour productivity. There are two common ways of measuring labour productivity. The first is *output per worker*. This is the most straightforward measure to calculate. All that is required is a measure of total output and employment.

A second measure is *output per hour worked*. This has the advantage that it is not influenced by the *number* of hours worked. So for an economy like the UK, with a very high percentage of part-time workers on the one hand, and long average hours worked by full-time employees on the other, such a measure would be more accurate in gauging worker efficiency.

International comparisons of labour productivity

Charts (a) and (b) show comparative productivity levels of various countries and the eurozone using GDP per hour worked. Chart (a) shows countries' productivity relative to the UK. GDP per hour worked has been consistently lower in the UK than in France, Germany and the USA. GDP per hour worked has been consistently higher in France, Germany and the USA than in the UK. For example, in 2021 compared with the UK it was 26 per cent higher in the USA, 15 per cent higher in Germany and 13 per cent higher in France. Recently, output per hour across the eurozone has risen to match that of the UK and in 2021 was 3 per cent higher. Japan is the exception having seen consistently lower GDP per hour worked than the UK. In 2021 it was 20 per cent less than in the UK.

A major explanation of lower productivity in the UK is the fact that for decades it has invested a smaller proportion of its national income than most other industrialised nations. Nevertheless, until 2007 the gap had been narrowing. This was because the level of UK productivity, although lower than in many other countries, was growing faster. This can be seen in chart (b). Part of the reason for this was the inflow of investment from abroad. However, since the global financial crisis, the gap has widened again. Some attribute this to the more flexible labour markets of the UK suppressing wage growth and thus reducing the pressure on employers to invest in more efficient technology.

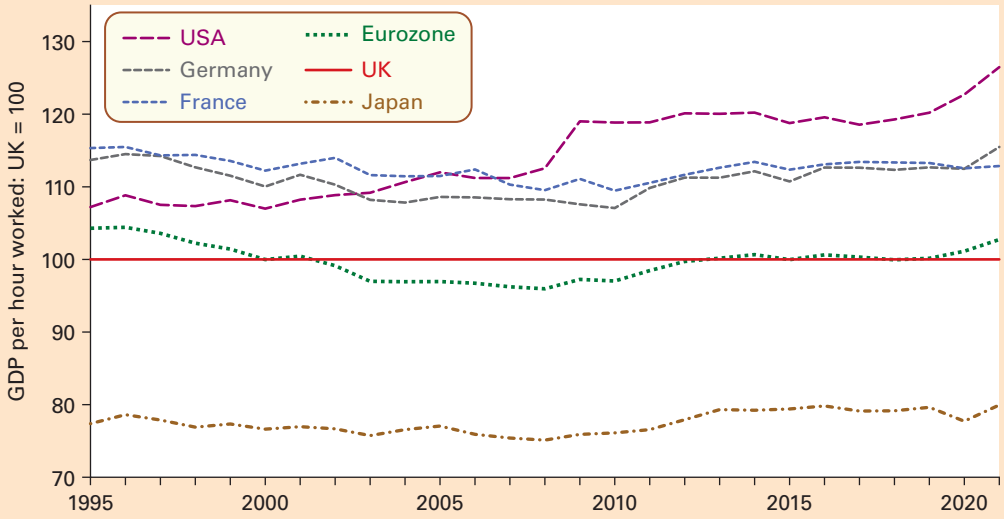
Chart (c) compares labour productivity across both measures. Workers in the USA and the UK on average work longer hours than those in France and Germany. Thus, whereas output *per hour worked* in the USA is around 10 to 12 per cent higher than in France and Germany, output *per person employed* in the USA is about 34 per cent higher than in France and 45 per cent higher than in Germany. As you can see, the evidence points to UK labour productivity being *lower* than that in the USA, France and Germany on both measures but higher than that in Japan.

In understanding the growth in labour productivity we need to focus on three factors: physical capital (see Box 10.3), human capital (see Box 10.4), and innovation and technological progress. The significance of these for the UK productivity gap is considered further in Case Study 10.3 on the student website.

1. *Identify some policies a government could pursue to stimulate labour productivity growth.*
2. *What could explain the differences in labour productivity between the five countries in chart (c), and why do the differences vary according to which of the two measures is used?*

Do a search to find productivity levels and growth rates in two other developed countries. Explain your findings in comparison to the countries in the charts in this box.

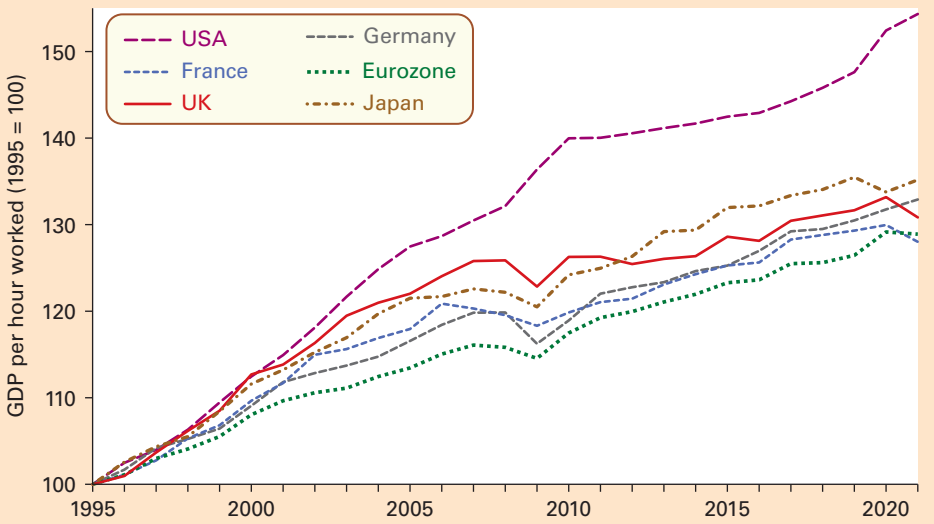
(a) Productivity in selected economies relative to the UK (GDP per hour worked)



Note: Figures are based on constant-price PPS GDP per hour worked

Source: Based on data in *OECDStat* (OECD, 2022)

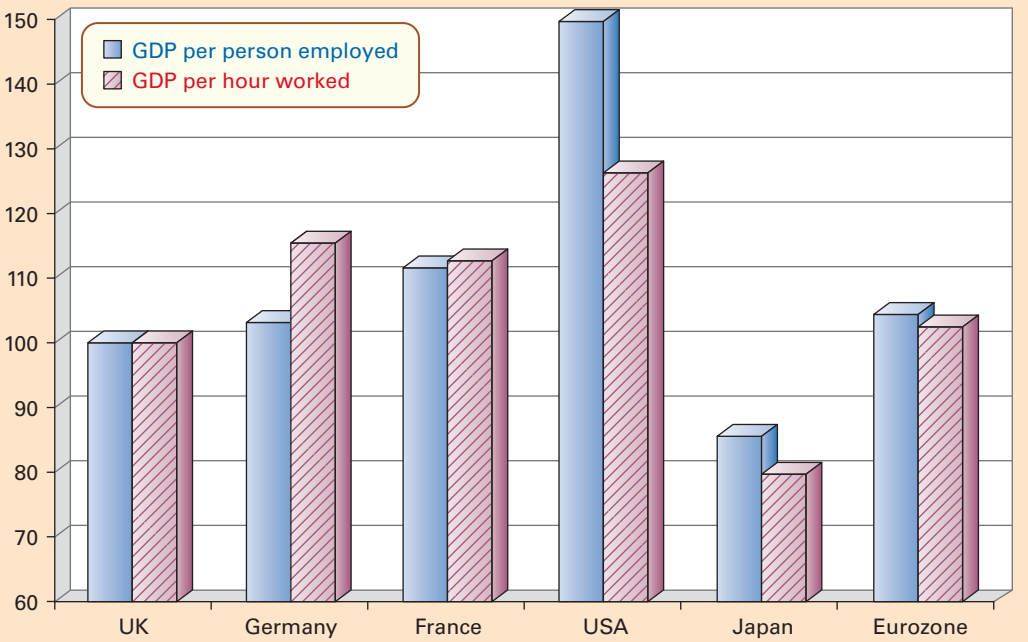
(b) Productivity in selected economies (GDP per hour worked, 1995 = 100)



Note: Figures are constant-price PPS GDP per hour worked

Source: Based on data in *OECDStat* (OECD, 2022), rebased by the authors

(c) Productivity in selected economies, 2021 (UK = 100)



Note: Figures are based on constant-price PPS GDP per hour worked/person employed

Source: Based on data in *OECDStat* (OECD, 2022)

BOX 10.3 GETTING INTENSIVE WITH PHYSICAL CAPITAL

Capital intensity and labour productivity

KI 34 p 287

In this box we take a look at two issues relating to physical capital. First, we consider what counts as capital in a country's national accounts. Second, we compare the growth of the physical capital stock per worker of a sample of developed economies and then see how this compares with their rates of growth in output per worker (labour productivity).

What is capital?

In a country's national accounts, physical capital consists of non-financial *fixed assets*. It does not include goods and services transformed or used up in the course of production; these are known as *intermediate goods and services*. Furthermore, it does not relate directly to the stock of human capital: the skills and attributes embodied in individuals that affect production (see Box 10.4).

A country's stock of fixed assets can be valued at its replacement cost, regardless of its age: this is its *gross* value. It can also be valued at its written-down value known as its *net* value. The net value takes into account the *consumption of capital* which occurs through wear and tear (depreciation) or when capital becomes naturally obsolescent.

The table shows that the estimated value of the *net* capital stock of the UK in 2021 was £4.8 trillion, 2.1 times the value of GDP. This, however, is considerably less than the 2020 estimate of *human* capital of £24 trillion, which was 11 times the size of GDP (see Box 10.4).

There are seven broad categories of fixed assets.

- The largest of these by value is *dwellings,* which includes houses, bungalows and flats. Residential housing yields rental incomes for landlords and, more generally, provides all of us with important consumption services, most notably shelter.
- The second largest component by value is *other buildings and structures.* This includes structures such as factories, schools and hospitals and the country's railway track.
- The third largest component by value is *ICT and other machinery, equipment and weapons systems.* It includes telecommunications equipment, computer hardware, office machinery and hardware as well as weapon systems equipment.

UK net capital stock, 2021

Type	£ billions	% of fixed assets	% of GDP
Dwellings	1931	40.3	83.3
Other buildings and structures	1624	33.9	70.1
ICT, other machinery, equipment & weapons systems	486	10.1	21.0
Intellectual property products	338	7.0	14.6
Land improvements	268	5.6	11.6
Transport equipment	139	2.9	6.0
Cultivated biological resources	8	0.2	0.3
All fixed assets	4795	100.0	206.9

Sources: Based on data from *Capital Stocks, Consumption of Fixed Capital in the UK dataset* and series YBHA (National Statistics)

- The fourth largest is *intellectual property products.* This includes computer software, original works of literature or art and research and development. The fifth largest category is *land improvements*: clearance and excavation that increases the value of land. The next largest is *transport equipment,* which includes equipment for moving people and objects such as lorries for haulage, buses, railway rolling stock and civil aircraft.
- The smallest component by value is *cultivated biological resources.* This includes livestock for breeding, vineyards, orchards and forests.

How does capital grow? An international comparison

A key measure of the amount of capital being used in an economy is the amount of capital per worker (capital intensity). In the chart we plot the average annual rate of growth of real GDP per worker since 1961 (y-axis) against the average annual rate of growth of capital per worker (x-axis) in a selection of developed countries. Therefore,

known as **capital deepening**. If the capital to labour ratio (K/L) were to fall, rather than to rise, then **capital shallowing** occurs.

Capital deepening and the resulting increase in capital per worker will generally increase output. In other words, the more equipment that is used by people at work, the more they are likely to produce. But to increase capital requires investment, and that investment requires resources – resources that could have been used for producing consumer goods. Thus more investment over the longer term requires more saving.

TC 1 p 7

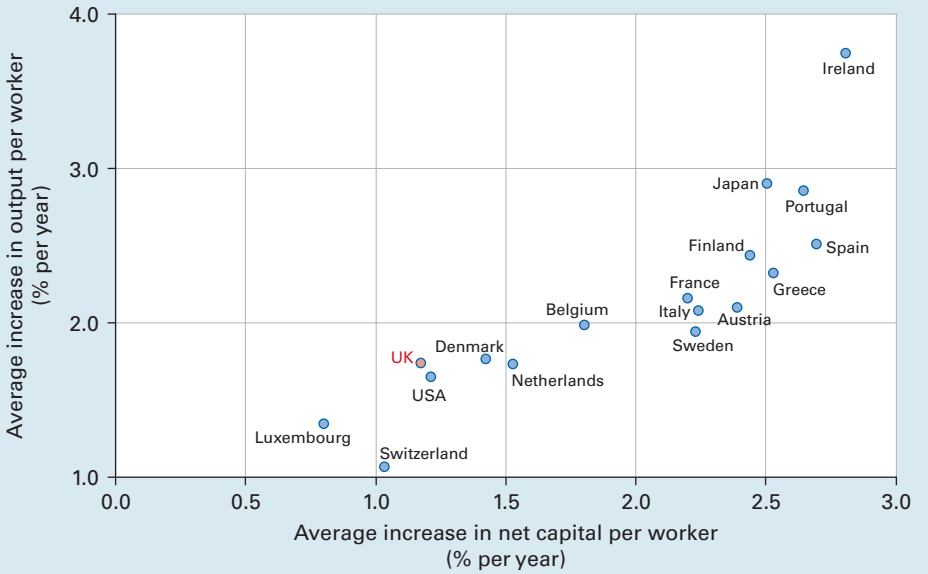
The growth of a country's capital stock is therefore closely related to levels of investment. Figure 10.10 illustrates this point for the UK. It shows the marked slowdown in the growth of the country's real net capital stock (the market value of the capital stock after accounting for the depreciation of assets) following the global financial crisis and COVID-19 pandemic. This is mirrored by a fall in the levels of investment relative to the size of the net capital stock.

KI 23 p 173

Box 10.3 considers the different types of physical capital as recorded in countries' national accounts. It

EXPLORING ECONOMICS

Growth in capital and output per worker, 1961–2023



Note: Figures from 2021 are forecasts; output and net capital are measured at constant prices

Source: Based on data from *AMECO Database* (European Commission)

the chart is plotting the increase in labour productivity against the increase in capital intensity.

For each country we observe an increase in capital intensity, although the rates of capital deepening differ quite significantly. The data show that the UK ranks relatively lowly in terms of capital deepening. In the UK the capital stock per worker has grown at an average annual real rate of 1.2 per cent. This compares with, for example, Japan, where the rate is 2.5 per cent, or Spain, where it is 2.7 per cent or Ireland, where it is 2.8 per cent.

We would expect that the more rapid the rate of capital accumulation, the faster the growth of output per worker (labour productivity). This is largely borne out in the chart. However, while there is a strong statistical association between capital deepening and economic growth, there are other factors that impact on long-term growth. Three of these are technological progress, *human* capital and the efficiency with which capital is deployed. We examine these later in this chapter.

1. *How does capital accumulation differ from capital deepening?*
2. *Could the composition as well as the size of a country's capital be important for its long-run economic growth?*

From the AMECO database download data on the net capital stock at constant prices per person employed. Then, for a sample of up to five countries of your choice, plot a time series chart showing the year-to-year rates of growth across time. Finally, comprise a short briefing note summarising the patterns in your chart.

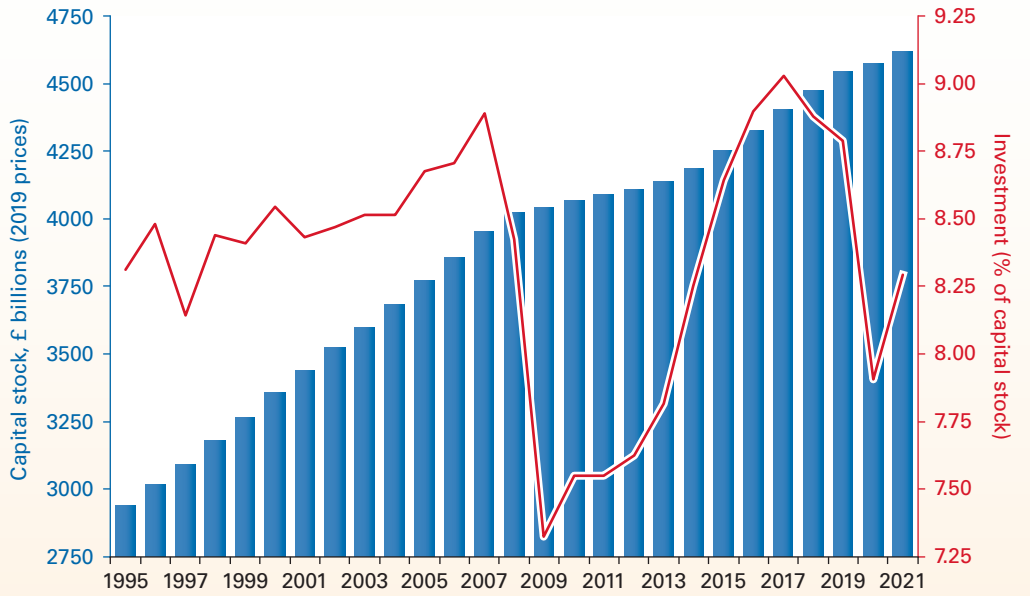
also looks at the extent to which rates of capital accumulation vary across developed economies.

A simple model of growth

The **Harrod–Domar growth model** (named after the two economists, Sir Roy Harrod and Evsey Domar, who independently developed the model) assumes that the rate of growth in real national income (g) via capital accumulation depends on two things.

- The marginal capital/output ratio (k). This is the amount of extra capital (ΔK) divided by the extra annual output that it produces (ΔY). Thus $k = \Delta K/\Delta Y$. The lower the value of k, the higher is the productivity of capital (i.e. the less extra capital you need to produce extra output).
- The proportion of national income that is invested (i), which, assuming that all saving is invested, will equal the proportion of national income that is saved (s).

Figure 10.10 UK net capital stock and investment



Note: Investment is Gross Fixed Capital Formation (whole economy) at constant 2019 prices

Source: Based on data from series *MLR3* and *NPQT* (National Statistics)

With a constant labour force, the Harrod–Domar formula for growth becomes:

$$g = i/k (\text{or } g = s/k)$$

Thus if 20 per cent of national income went in new investment ($i = 20$ per cent), and if each £1 of new investment yielded 25p of extra income per year ($k = 4$), then the growth rate would be 5 per cent.

A simple example will demonstrate this. If national income is £100 billion, then £20 billion will be invested ($i = 20\%$). This will lead to extra annual output of £5 billion ($k = 4$). Thus national income grows to £105 billion: a growth of 5 per cent.

However, we need to make three qualifications to this simple model to gain a better understanding of how increases in physical capital affect economic growth.

- *Diminishing returns to capital*. The first is that as capital per worker increases, so diminishing returns to capital are likely to set in. For example, if, in an office, you start equipping workers with PCs, at first output will increase very rapidly. But as more and more workers have their own PC rather than having to share, so the rate of increase in output slows down. When everyone has their own, output is likely to be at a maximum. Any additional PCs (of the same specification) will remain unused. Thus, for a given workforce, as the capital stock increases, so the marginal capital/output ratio (k) will rise and the growth rate will fall.

KI 17 p 95

- *Required investment for replacement purposes*. The second qualification is that the larger the capital stock, the greater the amount of investment that will be required for replacement purposes. This means that a smaller proportion of a given amount of investment is available for increasing the size of the capital stock at higher capital stock levels. Instead, more and more replacement investment is needed simply to maintain existing stock levels.

KI 23 p 173

- *Required investment for new workers*. The third qualification is to allow for changes in the size of the workforce. For capital deepening to occur (an increase in the capital to labour ratio), capital must increase by a larger percentage than any increase in workers.

KI 23 p 173

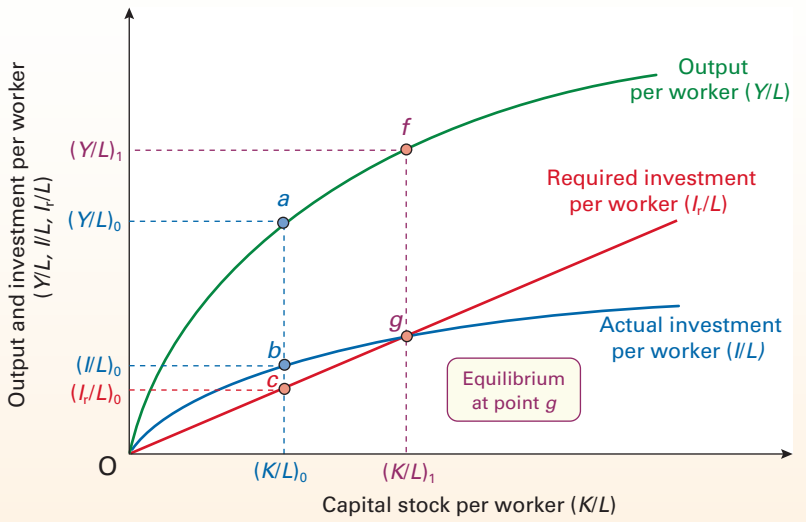
The neoclassical growth model

Constructing the neoclassical growth diagram

Let us now incorporate these qualifications into a model of growth. This is known as the **neoclassical** or **'Solow' growth model**, after the MIT economics professor and Nobel Prize winner, Robert Solow. This model is illustrated in Figure 10.11. The size of the capital stock per worker (K/L) is measured on the horizontal axis; the level of output per worker (Y/L), i.e. labour productivity, is measured on the vertical axis.

So how have we taken the three qualifications into account?

Figure 10.11 Steady-state output



KI 17 p 95 First, consider the assumption of diminishing returns to capital. The Y/L curve shows that as the capital stock per worker increases, so output per worker increases. The curve is commonly referred to as the **economy's production function** or the **aggregate production function**. But, because of diminishing returns to capital the economy's production function gets less and less steep.

What will be the effect on investment per worker? Increased output per worker will mean increased saving and hence increased investment per worker. This is shown by the *actual* investment per worker curve (the I/L curve), which has the same general shape as the Y/L curve because saving is assumed to be a given proportion of GDP. The higher the share of saving and investment in GDP, the steeper will be this line.

KI 23 p 173

Second, we consider the amount of investment required to replace existing capital that is wearing out or becoming obsolete. For a given annual rate of depreciation of capital, the bigger the capital stock, the larger the amount of replacement investment that will be required. The *required* investment per worker curve (the I_r/L curve) line incorporates this required level of replacement investment. It is upward sloping because replacement investment will increase the greater the capital stock per worker. The greater the rate of depreciation, the steeper this line will be.

KI 23 p 173

Third, we consider the growth in the workforce. It is assumed that the number of workers grows at a constant rate per annum. To maintain a given capital/labour ratio, new workers will need to be equipped with capital. The required investment to do this is also incorporated in the required investment per worker curve (I_r/L). The higher the rate of growth in the workforce, the more investment will be required to maintain a given capital/labour ratio and hence the steeper the I_r/L curve.

The I_r/L line will rotate upwards around the origin if either the rate of depreciation or the rate of growth of the workforce increases. This is because at any level of capital per worker more investment would be required for either greater amounts of depreciation or for equipping larger numbers of additional workers.

KI 23 p 173

Pause for thought

In the absence of growth in the workforce what does the required investment per worker curve capture?

Definitions

Neoclassical analysis The analysis of market economics where it is assumed that individuals and firms are self-interested rational maximisers.

Solow growth model A model which explains economic growth in terms of the effects on the capital stock and output of a change in investment.

Economy's production function (or **aggregate production function**) The relationship between the economy's output and output per worker (labour productivity), holding the level of human capital and the state of technology constant.

The process to a 'steady state'

Where is the economy's long-run equilibrium point in the model?

Assume initially that the size of the capital stock is $(K/L)_0$ as shown in Figure 10.11. This will generate an output per worker of $(Y/L)_0$ (point *a*). This output, in turn, will generate saving and (actual) investment per worker of $(I/L)_0$, but of this, $(I_r/L)_0$ will have to be used for replacement purposes and for equipping additional workers. The difference $(b - c)$ will be available for capital deepening: to increase the size of the capital stock per worker.

KI 23 p 173

The capital stock per worker will thus increase up to $(K/L)_1$ (point *g*). At this point, no further capital deepening will occur (only **capital widening**): all investment is merely replacing depreciating capital and equipping new workers to maintain the existing capital intensity. Output per worker will therefore cease growing. $(K/L)_1$ therefore represents the steady-state level of capital per worker and $(Y/L)_1$ the **steady-state level of real national income per worker**.

At the steady state, as the names infers, capital *per worker* (K/L) and output *per worker* (Y/L) – and hence labour productivity – have ceased growing. However, the levels of capital (K) and output (Y) are growing, but merely at the rate at which the labour force (L) is growing. This must be the case for the capital per worker (K/L) and output per worker (Y/L) ratios to be constant.

Pause for thought

What is happening to the general standard of living at the steady state?

An increase in the saving rate

In the simple model, $g = s/k$, an increase in the saving rate (s) will increase the growth rate (g). When we take into account diminishing returns to capital, depreciation and a growth in the number of workers, however, an increase in the saving rate will lead to only a temporary increase in output per worker. Therefore, there is no long-term growth in output per worker at all!

KI 17 p 95

If the saving rate increases, the investment curve will shift upwards. This is shown by a shift from $(I/L)_1$ to $(I/L)_2$ in Figure 10.12. Investment is now above that which is necessary to maintain the capital stock at $(K/L)_1$. Capital intensity will rise, therefore, and so will national income per worker. But this growth is only temporary. Once the capital stock has risen to $(K/L)_2$, all the new higher level of investment will be absorbed in replacing capital and equipping additional workers. Required investment equals actual investment at point *n*. Real national income per worker stops rising. $(Y/L)_2$ represents the new steady-state real national income per worker.

KI 23 p 173

Does this mean, therefore, that there is no long-term gain from an increase in the saving rate? There *is* a gain, to the extent that real national income per worker is now higher, and this higher average income will be received not just once, but every year from now on as long as the saving rate remains at the new higher level. However, what should be clear from the above analysis is that, without technological progress or some other means of increasing the average output of workers, long-term economic *growth* cannot be sustained.

Pause for thought

If there were a gradual increase in the saving rate over time, would this lead to sustained economic growth even without technological progress?

Definitions

Capital widening A proportionate rise in both capital and labour. There is more capital but no capital deepening.

Steady-state real national income per worker The long-run equilibrium level of real national income (real GDP) per worker. The level at which all investment is used merely to maintain the existing capital stock per worker at its current level.

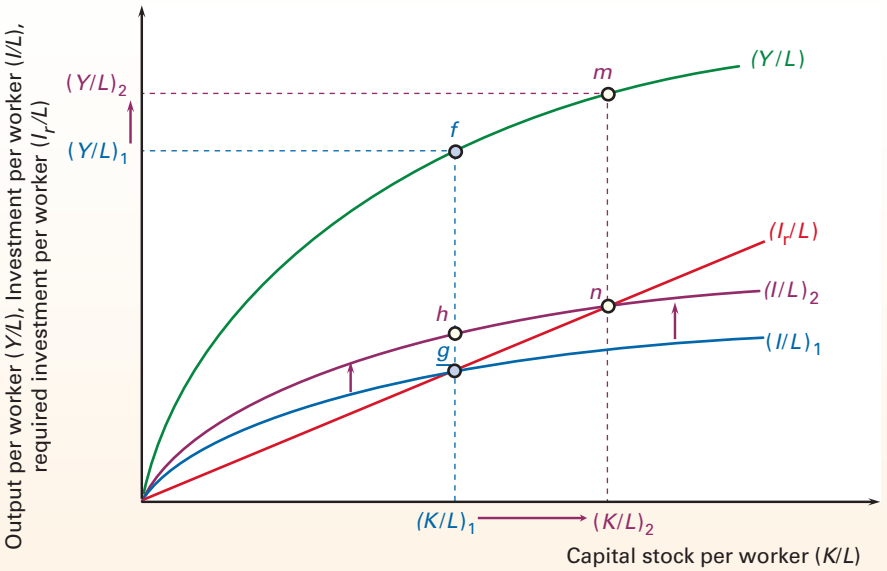
Human capital and education

The analysis of Figures 10.10 and 10.11 need not be confined to the stock of physical capital: machines, buildings, tools, etc. It can also apply to human capital. Human capital, as we saw in Section 10.4, refers to the skills and expertise of workers that have, in part, been acquired through education and training. If part of saving is used for investment in education and training, then the productivity of workers will rise, and so will output.

In Figures 10.10 and 10.11, therefore, the definition of capital can be extended, such that the horizontal axis measures both physical and human capital. An increase in either has the effect of increasing the steady-state level of national income per worker.

Box 10.4 looks at recent estimates of the stock of human capital in the UK.

Figure 10.12 Effect of an increase in the rate of saving and investment



Recap

1. Capital deepening occurs when the economy's capital intensity (K/L) rises. This is important for long-term economic growth as it enables output per worker (Y/L) to rise. The rate of capital deepening is crucial in determining the long-term rate of growth labour productivity and hence in living standards.
2. The Harrod–Domar model of economic growth shows that an increased saving rate will lead to higher investment and hence to an increase in the capital stock. This, in turn, will lead to a higher level of real national income.
3. The neoclassical growth model analyses the process of capital accumulation. It makes three important qualifications. First, the economy's production function is characterised by diminishing returns to capital.; second, as the capital per worker ratio rises more investment is needed for replacement purposes; and third more investment will be needed to equip any new workers with the same level of capital as existing workers.
4. The neoclassical model shows how capital deepening eventually ceases if actual levels of investment are absorbed by all the extra investment required to maintain existing levels of capital intensity. Hence, national income per worker and capital per worker stop rising. A steady-state level of real national income per worker has been achieved.
5. At the steady-state, output per worker (Y/L) is constant. Therefore, an economy's output level (Y) is growing only at the rate at which the labour force is growing. In the absence of any growth in the labour force, output growth is also zero.
6. An increased saving rate will therefore lead only to a rise in the *level* of output per worker. A new, higher steady-state level of real national income per worker is reached but growth has again ceased.

10.5 ECONOMIC GROWTH WITH TECHNOLOGICAL PROGRESS

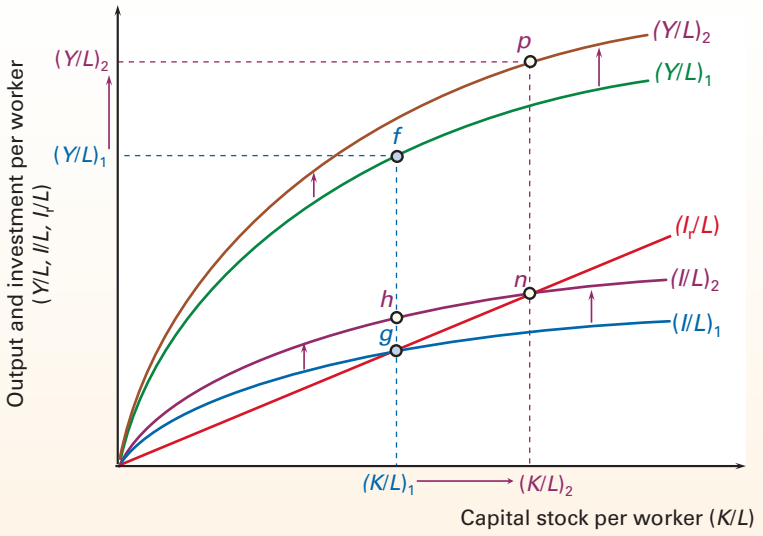
Can technology help to sustain rising labour productivity?

Technological progress and the neoclassical model

TC 15 p 287

Technological progress has the effect of increasing the output from a given amount of investment. This is shown in Figure 10.13. The initial investment and output per worker curves are $(I/L)_1$ and $(Y/L)_1$; steady-state income is at a level of $(Y/L)_1$ (point f). A technological advance has the effect of shifting the (Y/L) line upwards, say to $(Y/L)_2$. This higher aggregate production function leads to a higher investment per worker curve (for a given rate of saving). This is shown by curve $(I/L)_2$. The new long-term equilibrium capital stock is thus $(K/L)_2$, and the new steady-state level of income is $(Y/L)_2$ (point p).

Figure 10.13 Effect of a technological advance



If there is a 'one-off' technological advance, the effect is the one we have just illustrated. Real national income rises to a higher level, but does not go on rising once the new steady-state level has been reached. But technological progress marches on over time. New inventions are made; new processes are discovered; old ones are improved. In terms of Figure 10.13, the aggregate production function *goes on* shifting upwards over time, as does the (I/L) curve too. The faster the rate of technological progress, the faster the aggregate production function will shift upwards and the higher the rate of growth of output per worker (real national income per worker) will be.

Pause for thought

What impact does the rate of technological progress have on the growth of labour productivity?

Endogenous growth theory

Continual technological advance is essential to sustaining capital deepening and growth in output per worker and labour productivity. Hence, an increase in technological progress is essential if a country wants to achieve faster rates of growth in the long term. But is this purely in the lap of the scientists and engineers?

TC 15 p 287

In the Solow growth model that we have been considering up to now, this is indeed the type of assumption made. In other words, technological progress is simply a 'given': it is exogenously determined. Thus, although the neoclassical model identifies the importance of technological progress for enduring growth, it does not offer governments actual policy prescriptions.

What can be done to speed up the rate of innovation? Can governments adopt policies that encourage scientific breakthroughs and technological developments? **Endogenous growth models** attempt to answer such questions by incorporating technological advancement *within* them.

Endogenous growth models stress the importance of research and development, education and training, and fostering innovation. Hence, policy makers are interested in appropriate policies that might, for example, make their country a world leader in innovation and technological advancement, especially if this helps to sustain higher long-run rates of economic growth, and so higher standards of living.

In such models, a major determinant of technological progress is the size and composition of the capital stock.

Definition

Endogenous growth models Models where the rate of economic growth depends on the rate of technological progress and diffusion, both of which depend on size of the capital stock and the capital goods industries, and also on institutions, incentives and the role of government.

As economies accumulate capital, they are likely to devote more resources to the development and maintenance of capital goods. In other words, they are likely to have a larger sector devoted to producing and developing capital goods. Therefore, the very process of capital accumulation can raise the rate of technological progress and enable even further capital accumulation.

Endogenous growth models describe how the interaction between capital accumulation and technological progress generates a virtuous circle. Consequently, rather than looking at capital accumulation and technological progress as purely separate sources of long-term growth, the two are arguably interdependent.

Pause for thought

What is 'endogenous' about endogenous growth theory?

Investment in research and development can be encouraged through the use of patents and copyrights. These provide some protection to firms, enabling them to capture more of the benefits from their own ideas and thus providing them with an incentive to create and innovate. Furthermore, the striving for profit or the pursuit of competitive advantage over rivals, either through the design of innovative products and services or more cost-efficient production processes, are incentives that can drive this innovation and creativity.

But there are limits to the ability of firms to exclude other firms from prospering from their own ideas, such as the development of products, processes and people. The virtuous circle is thus reinforced by *externalities*: the spill-over of ideas from one firm to another. New ideas cannot be put back into the metaphorical bottle once its lid is off.

KI 28 p 196

Pause for thought

Are the externalities in endogenous growth models externalities in consumption or production?

A model of endogenous technological progress

Endogenous growth models argue two things. The first is that technological progress is *dependent* on various economic factors such as the rate of investment in research and development. This could be included as an element in the investment (I) term, i.e.

$$I = I_n + I_c$$

where I_n is investment in research and development of new technology (it could also include investment in training) and I_c is investment in capital that uses current technology. The greater the value of I_n/I_c, the faster both the aggregate production function (the Y/L curve) and the investment per worker curve (the I/L curve) will shift upwards in Figure 10.13. Any policy, then, that increases the proportion of national income being devoted to R&D and training, will increase the long-run rate of economic growth.

The second element is the responsiveness of national income to new technologies. This will depend in part on the extent to which innovations spill over to other firms, which duplicate or build on them, thereby adding to the increase in national income (ΔY). But, it will also depend on the ability of people to exploit the opportunities afforded by these potential spillovers.

The greater the value of $\Delta Y/I_n$, the steeper the *aggregate production function* in Figure 10.13 will be and the greater the rate of economic growth will be.

The values of I_n and $\Delta Y/I_n$ are therefore thought to depend on structural and institutional factors within the economy and on the role of government. These include:

- attitudes of entrepreneurs and businesses, such as their inclination to take risk;
- the willingness of financial institutions to lend to support investment opportunities;
- tax incentives and government grants, for instance support for R&D;
- a research infrastructure (e.g. laboratories and the number and skills of researchers);
- the degree of competition within industries;
- the incentives to develop new products and services and/or to reduce costs;
- the magnitude of external spillovers from the generation of new products, processes and techniques;
- the stock of human capital.

Pause for thought

List a series of factors that you think could influence the significance of spillovers from investment.

The economy's production function

As we have seen, endogenous growth models can explain how the economy's production function shifts upwards over time. The process of capital accumulation creates the conditions for technological progress.

TC 15 p 287

In some cases, the production function may also become *steeper*. In other words, a given rise in the capital stock causes a larger rise in national income.

BOX 10.4 UK HUMAN CAPITAL

Estimating the capabilities of the labour force

The OECD (2001) defines human capital as 'the knowledge, skills, competencies and other attributes embodied in individuals or groups of individuals acquired during their life that facilitate the creation of personal, social and economic well-being'.[1] Hence, trends in human capital have implications for a range of economic-related issues, including economic growth, unemployment, life satisfaction, the inequality of income, wealth and opportunity and also for social cohesiveness. But, how we do we go about measuring human capital?

Measuring human capital

In estimating an individual's human capital, a common approach is to estimate the present value of an individual's *remaining lifetime labour income*. This can be done for representative individuals in categories defined by gender, age and educational attainment. An assumption is then made about the working life of individuals. In compiling the UK estimates it is assumed that the remaining lifetime labour income of individuals aged 65 and over is zero. Then an approach known as *backwards recursion* is applied.

Backwards recursion involves first estimating the remaining lifetime labour income of someone aged 64 with a particular gender, age and educational level. The remaining lifetime income in this case is simply their current annual labour income for the year from their 64th birthday. For someone aged 63 it is their current annual labour income for the year from their 63rd birthday plus the present value[2] of the remaining lifetime income of someone aged 64 with the same gender, age and educational level. This continues back to someone aged 16. In calculating the remaining lifetime labour income of representative individuals, account is also taken of the probability that their level of education attainment may rise and, with it, their expected future earnings.

Further working assumptions are necessary to complete the calculations. Two of the most important are that: the rate of labour productivity growth is 2 per cent per annum and the discount rate is 3.5 per cent per annum, as recommended by HM Treasury's Green Book (2003) when undertaking appraisal and evaluation studies in central government.

Two measures of the stock of human capital are estimated. The first is for *employed* human capital. It is based on estimating the lifetime labour income of those in employment. The second is *full* human capital. It includes the human capital of both the employed and unemployed. This assumes that the human capital of those currently unemployed should be valued at the remaining lifetime labour income of employed individuals with the same characteristics (gender, age and educational attainment). It ignores any so-called scarring effects from being unemployed, such as the depreciation of job-specific or transferable skills. Such effects are likely to increase the longer the duration of unemployment.

Estimates of human capital

The chart shows estimates of employed and full human capital in the UK since 2004. Both follow broadly similar patterns. Between 2004 and 2011, the stock of (full) human capital in the UK increased by an average of 1.30 per cent per annum (see chart). However, there was then a marked slowdown: between 2012 and 2019 it grew by only 0.54 per cent per annum. Nonetheless, by 2020 the constant-price estimates of human capital had risen to £23.8 trillion – up from £20.6 trillion in 2004 (both in 2020 prices). This was equivalent to 11 times the size of annual GDP.

Changes in the stock of human capital can be decomposed into several factors, including changes in educational attainment and population, changes in the gender balance and the impact of the age structure of society. Over the period since 2004, population growth and rising educational attainment, particularly the growth in

[1] *The Well-Being of Nations: The role of human and social capital*, Centre for Educational Research and Innovation, OECD (2001).

[2] 'Present value' in this case is the value in today's terms of income earned in the future. These incomes have to be reduced by the rate of interest that could have been earned if the income had been earned today instead of in the future. This process of reducing future incomes to present values is known as 'discounting'.

The reason is that the benefits of the output from investment are not just captured by the firms doing the investing. Rather, some spill over to other firms. For example, firms may be able to duplicate or develop other firms' ideas. Consequently, these spillovers can positively impact on the overall marginal product of capital.

Policy implications

If there is a virtuous circle arising from firms investing and innovating, how are governments best placed to encourage it? Many economists argue that this requires supply-side policies: policies that impact directly on aggregate supply. Examples include policies to influence research and development, education and training, industrial organisation and work practices.

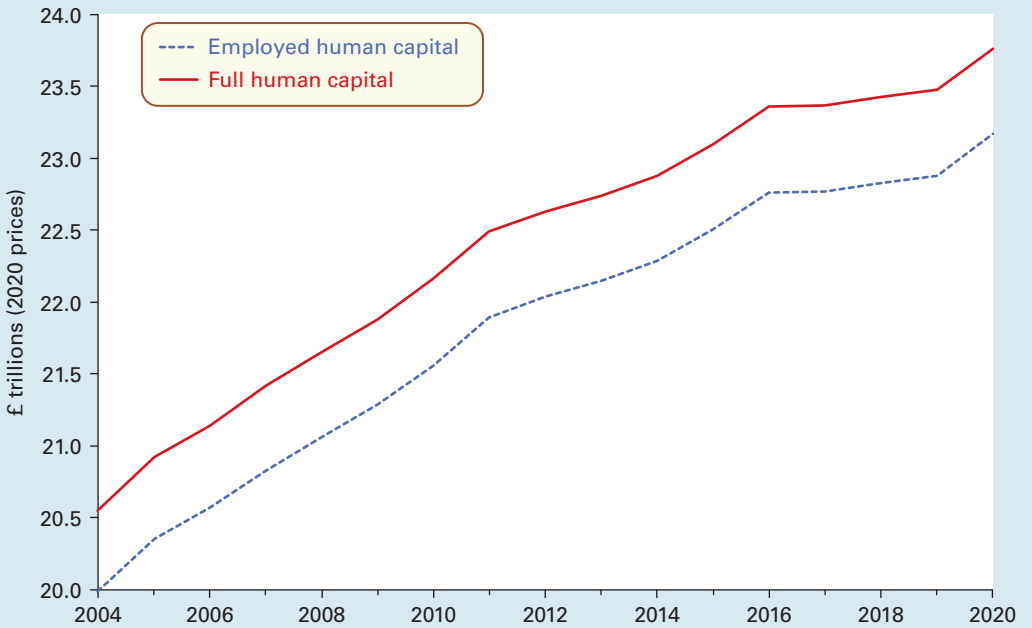
TC 15 p 287

There is less agreement, however, as to whether these policies should focus on delivering market solutions or involve greater state intervention. We return to these themes in Section 13.4, which looks at a range of supply-side policies.

But encouraging investment does not just depend on effective supply-side policies. It also depends on the stability of the macroeconomic environment. It is not easy for businesses to plan ahead in times of great economic uncertainty. In the years following the global

EXPLORING ECONOMICS

Employed and full human capital (2020 prices)



Source: Based on data in *Human Capital Estimates in the UK 2004 to 2020* (Office for National Statistics, April 2022)

numbers with degree-equivalent qualifications, have had the most significantly positive effects on human capital.

In more recent times, however, their impact has lessened, while, at the same time, an ageing population has reduced the growth in the stock of human capital. It will be interesting to see in due course whether the disruptions to education during the pandemic have had any long-term detrimental effect on human capital growth.

Despite this, degrees remain an important influence on human capital. An analysis of the distribution of human capital by educational attainment in 2020 found that 43.4 per cent of UK employed human capital was embodied in those who had a degree (or equivalent).

1. *In what ways are human capital and physical capital complementary?*
2. *Other than by educational attainment, in what ways might we wish to analyse the distribution of human capital?*

Using the human capital statistics dataset from the ONS, calculate the percentage shares of human capital originating from the different regions and countries of the UK. Summarise your findings in a short briefing note.

financial crisis of 2007–8, uncertainty led to lower levels of investment and innovation in many countries, causing productivity and potential output to grow more slowly. Similar concerns for long-term growth arose following the COVID-19 pandemic and the turbulent macroeconomic environment that characterised the early 2020s.

By contrast, the less the volatility in output (real GDP), the greater the confidence of business to innovate and invest will be and the higher the growth in potential output will be.

The idea that economic shocks can have persistent or enduring effects on the path of potential output has become an accepted feature of many macroeconomic models. Whatever the source of these shocks, it is important to understand the mechanisms by which they are transmitted through the economy and their significance for longer-term growth.

Some economists, however, argue that, given the inherent volatility of economies, governments *should* be more proactive. For instance, fluctuations in credit and financial conditions can have a marked effect on flows of investment, including spending on research and development. They argue that governments have an important role to play in helping to stabilise aggregate demand so as to support and encourage firms to invest and thereby increase potential output.

Recap

1. Technological progress enables capital accumulation. The more rapid the pace of technological progress is the higher the rate of growth in labour productivity will be and hence the higher the long-term rate of economic growth will be.
2. Endogenous growth models show how capital accumulation and technological progress interact. Hence, capital accumulation and technological progress are interdependent sources of long-term growth.
3. Endogenous growth theories suggest that the rate of technological progress and its rate of diffusion depend on economic institutions and incentives. Supply-side policy could be used to alter these.

QUESTIONS

1. Why does the aggregate demand curve slope downwards? What factors would cause it to move to the left?
2. In what way will the nature of aggregate supply influence the effect of a change in aggregate demand on prices and real national income?
3. What shape do you think the aggregate supply curve would be at the current output if the economy was in a deep recession?
4. What shape of aggregate supply curve is assumed by the simple Keynesian demand-driven model of the economy introduced in Chapter 9? Under what circumstances is this shape likely to be a true reflection of the aggregate supply curve?
5. Referring to Figure 10.1, assume that the price level is currently above the equilibrium. Explain how the price level would return to its equilibrium level.
6. What are the determinants of long-run economic growth? Is long-run economic growth sustainable without technological progress?
7. What is meant by labour productivity? What are the sources of growth in labour productivity?
8. What distinguishes capital accumulation from capital deepening? Can capital intensity fall despite capital accumulation?
9. What determines the rate of depreciation? What would happen if the rate of depreciation fell?
10. Explain the shapes of the I/L and Y/L curves in the neoclassical model of growth. What happens to these curves if the share of investment in national income rises?
11. Describe the process of capital accumulation that results in an economy reaching a steady state in the neoclassical growth model.
12. At the steady state in the neoclassical growth model what is happening to the levels of output and output per worker? Under what circumstances would they be behaving in the same way?
13. What is the significance of the term 'endogenous' in endogenous growth theory? What, according to this theory, determines the long-run rate of economic growth?
14. What role do externalities play in endogenous growth theory?
15. What policy prescriptions do the neoclassical and endogenous growth theories offer policy makers looking to raise their country's long-run growth rate?
16. For what possible reasons may one country experience a persistently faster rate of economic growth than another?
17. How might significant economic shocks, such as the global financial crisis and the COVID-19 pandemic, have 'scarring effects' on an economy's potential output?

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

10.1 Introducing theories of economic growth. An overview of classical and more modern theories of growth.

10.2 Growth accounting. This case study identifies factors that contribute to economic growth and shows how their contribution can be measured.

10.3 Productivity performance and the UK economy. A detailed examination of how the UK's productivity compares with that in other countries.

10.4 The USA: is it a 'new economy'? An examination of whether US productivity increases are likely to be sustained.

10.5 Technology and economic change. How to get the benefits from technological advance.

10.6 UK industrial performance. This examines why the UK has had a poorer investment record than many other industrial countries and why it has suffered a process of 'deindustrialisation'.

10.7 The R&D Scoreboard. An international comparison of spending by companies on research and development.

10.8 Globalisation and the new economy. Does globalisation bring economic success?

WEB APPENDIX

10.1 Deriving a price index for the economy. This explores how the process of chain linking is used to derive price indices for the components of aggregate demand and for the economy as a whole.

Chapter 11

The financial system, money and interest rates

In this chapter we are going to look at the important roles that money and financial institutions play in the economy. Changes in the behaviour of financial institutions and in the amount of money can have a powerful effect on the macroeconomic environment, as was demonstrated starkly by the financial crisis that developed during the late 2000s. It is, therefore, important to understand the behaviour of the financial system and the means by which its actions transmit through the economy to affect all the major macroeconomic variables, such as inflation, unemployment, economic growth, exchange rates and the balance of payments.

In addressing how the financial system influences economic activity, we will analyse how the supply of money and the demand for money between them determine the *rate of interest*. We will see that this has a crucial impact on aggregate demand and therefore the performance of the economy. Furthermore, many aspects of economic activity are dependent on the availability of money.

The chapter begins by defining what is meant by money and examining its functions. Then in Sections 11.2 and 11.3 we look at the operation of the financial sector of the economy and its role in determining the supply of money. It is here that we consider the possible causes of the financial crisis, its impact on financial institutions and some of the responses by central banks to the problems faced by financial institutions and, as a result, the economy.

We then turn to look at the demand for money. What we are asking is: how much of people's assets do they want to hold in the form of money?

Next, we put money supply and demand together to see how free-market interest rates are determined. Finally, we will also see how changes in money supply and/or interest rates affect aggregate demand and the level of activity in the economy. By using a framework known as the 'quantity theory of money', we see how monetary changes can have important, but possibly uncertain effects on spending and national output.

After studying this chapter, you should be able to answer the following questions:

- What are the functions of money?
- What determines the amount of money in the economy? What causes it to grow and what is the role of banks in this process?
- Why do central banks, such as the Bank of England and the European Central Bank, play a crucial role in the functioning of economies?
- What is the relationship between money and interest rates? What is the role of various financial institutions in this relationship?
- How will a change in the money supply affect the level of aggregate demand? How will this, in turn, affect the level of real GDP?

11.1 THE MEANING AND FUNCTIONS OF MONEY

What is this thing called 'money'?

Before going any further, we must define precisely what we mean by 'money' – not as easy a task as it sounds. Money is more than just notes and coin. In fact, the main component of a country's money supply is not cash, but deposits in banks and other financial institutions. Only a very small proportion of these deposits are kept by the banks in their safes or tills in the form of cash. The bulk of the deposits appear merely as bookkeeping entries in the banks' accounts.

KI 32 p 239

This may sound very worrying. Will a bank have enough cash to meet its customers' demands? The answer in the vast majority of cases is yes. Only a small fraction of a bank's total deposits will be withdrawn at any one time, and banks always seek to ensure that they have the ability to meet their customers' demands.

The chances of banks running out of cash are very low indeed. The only circumstance where this could become possible is if people lost confidence in a bank and started to withdraw money in what is known as a 'run on the bank'. This happened with the Northern Rock Bank in September 2007. But in these circumstances the central bank or government would intervene to protect people's deposits by making more cash available to the bank or, in the last resort, by nationalising the bank (as happened with Northern Rock in February 2008).

KI 15 p 80

What is more, the bulk of all but very small transactions are not conducted in cash at all. By the use of debit cards, credit cards and cheques, most money is simply transferred from the purchaser's to the seller's bank account without the need for first withdrawing it in cash.

What items should be included in the definition of money? To answer this, we need to identify the *functions* of money.

The functions of money

The main purpose of money is for buying and selling goods, services and assets: i.e. as a medium of exchange. It also has two other important functions. Let us examine each in turn.

A medium of exchange

In a subsistence economy where individuals make their own clothes, grow their own food, provide their own entertainment, etc., people do not need money. If people want to exchange any goods, they will do so by barter. In other words, they will do swaps with other people.

The complexities of a modern developed economy, however, make barter totally impractical for most purposes (see Case Study 11.1 on the student website). What is necessary is a **medium of exchange** that is generally acceptable as a means of payment for goods and services and as a means of payment for labour and other factor services. 'Money' is any such medium.

To be a suitable physical means of exchange, money must be light enough to carry around, must come in a number of denominations, large and small, and must not be easy to forge. Alternatively, money must be in a form that enables it to be transferred indirectly through some acceptable mechanism. For example, money in the form of bookkeeping entries in bank accounts can be transferred from one account to another by the use of mechanisms such as debit cards and direct debits.

A means of storing wealth

People need a means whereby the fruits of *today's* labour can be used to purchase goods and services in the *future*. People need to be able to store their wealth: they want a means of saving. Money is one such medium in which to hold wealth. It can be saved.

A means of evaluation

Money allows the value of goods, services and assets to be compared. The value of goods is expressed in terms of prices, and prices are expressed in money terms. Money also allows dissimilar things, such as a person's wealth or a company's assets, to be added up. Similarly, a country's GDP is expressed in money terms. Money thus serves as a 'unit of account'.

A means of establishing the value of future claims and payments

People often want to agree today the price of some future payment. For example, workers and managers will want to agree the wage rate for the coming year. Firms will want to sign contracts with their suppliers specifying the price of raw materials and other supplies. Money prices are the most convenient means of measuring future claims.

Definition

Medium of exchange Something that is acceptable in exchange for goods and services.

Pause for thought

Why may money prices give a poor indication of the value of goods and services?

What should count as money?

What items, then, should be included in the definition of money? Unfortunately, there is no sharp borderline between money and non-money.

Cash (notes and coin) obviously counts as money. It readily meets all the functions of money. Goods (fridges, cars and cabbages) do not count as money. But what about various financial assets such as deposits in savings accounts and stocks and shares? Do they count as money? The answer is: it depends on how narrowly money is defined.

Countries thus use several different measures of money supply. All include cash, but they vary according to what additional items are included. To understand their significance and the ways in which money supply can be controlled, it is first necessary to look at the various types of account in which money can be held and at the various financial institutions involved.

Pause for thought

Why are debit cards not counted as money?

Recap

1. Money's main function is as a medium of exchange. In addition, it is a means of storing wealth, a means of evaluation and a means of establishing the value of future claims and payments.
2. What counts as money depends on how narrowly it is defined. All definitions include cash, but they vary according to what other financial assets are included.

11.2 THE FINANCIAL SYSTEM

Where do banks and other financial institutions fit in?

In order to understand the role of the financial sector in determining the supply of money, it is important to distinguish different types of financial institution. Each type has a distinct part to play in determining the size of the money supply.

The banking system

Retail and wholesale banking

By far the largest element of money supply is bank deposits. It is not surprising, then, that banks play an absolutely crucial role in the monetary system.

Banking can be divided into two main types: retail banking and wholesale banking. Most banks today conduct both types of business and are thus known as 'universal banks'.

Retail banking. **Retail banking** is the business conducted by the familiar High Street banks, such as Barclays, HSBC, Lloyds, Royal Bank of Scotland, NatWest (part of the RBS group) and Santander. They operate bank accounts for individuals and businesses, attracting deposits and granting loans at published rates of interest.

Wholesale banking. The other major type of banking is **wholesale banking**. This involves receiving large deposits from and making large loans to companies or other banks and financial institutions; these are known as wholesale deposits and loans.

As far as companies are concerned, these may be for short periods of time to account for the non-matching of a firm's payments and receipts from its business. They may be for longer periods of time, for various investment purposes. Because wholesale deposits and loans often involve very large sums of money, banks compete against each other for them

Definitions

Retail banking Branch, telephone, postal and Internet banking for individuals and businesses at published rates of interest and charges. Retail banking involves the operation of extensive branch networks.

Wholesale banking Where banks deal in large-scale deposits and loans, mainly with companies and other banks and financial institutions. Interest rates and charges may be negotiable.

BOX 11.1 FINANCIAL INTERMEDIATION EXPLORING ECONOMICS

What is it that banks do?

Banks and other financial institutions are known as **financial intermediaries**. They all have the common function of providing a link between those who wish to lend and those who wish to borrow. In other words, they act as the mechanism whereby the supply of funds is matched to the demand for funds. In this process, they provide five important services.

Expert advice

Financial intermediaries can advise their customers on financial matters: on the best way of investing their funds and on alternative ways of obtaining finance. This should help to encourage the flow of savings and the efficient use of them.

Expertise in channelling funds

Financial intermediaries have the specialist knowledge to be able to channel funds to those areas that yield the highest return. They also have the expertise to assess risks and to refuse loans for projects considered too risky or to charge a risk premium to others. This all encourages the flow of saving as it gives savers the confidence that their savings will be secure and earn a good rate of interest. Financial intermediaries also help to ensure that projects that are potentially profitable will be able to obtain finance. They help to increase allocative efficiency (see Section 8.1).

KI 25 p 195

Maturity transformation

Many people and firms want to borrow money for long periods of time, and yet many depositors want to be able to withdraw their deposits on demand or at short notice. If people had to rely on borrowing directly from other people, there would be a problem here: the lenders would not be prepared to lend for a long enough period. If you had £100 000 of savings, would you be prepared to lend it to a friend to buy a flat if the friend was going to take 25 years to pay it back? Even if there was no risk whatsoever of your friend defaulting, most people would be totally unwilling to tie up their savings for so long.

This is where a bank or building society comes in. It borrows money from a vast number of small savers, who are able to withdraw their money on demand or at short notice. It then lends the money to house purchasers for a long period of time by granting mortgages (typically these are paid back over 20 to 30 years). This process whereby financial intermediaries lend for longer periods of time than they borrow for is known as **maturity transformation**. They can do this because with a large number of depositors it is highly unlikely that they would all want to withdraw their deposits at the same time. On any one day, although some people will be withdrawing money, others will be making new deposits.

This does not mean that maturity transformation is without risks for financial institutions. Maturity transformation implies a maturity mismatch between the liabilities and assets on institutions' balance sheets. It needs managing. Box 11.2 considers the potential risks of maturity transformation for financial institutions and the wider economy.

Risk transformation

You may be unwilling to lend money directly to another person in case they do not pay up. You are unwilling to take the risk. Financial intermediaries, however, by lending to large numbers of people, are willing to risk the odd case of default. They can absorb the loss because of the interest they earn on all the other loans. This spreading of risks is known as **risk transformation**. What is more, financial intermediaries may have the expertise to be able to assess just how risky a loan is.

Transmitting payments

In addition to channelling funds from depositors to borrowers, certain financial institutions have another important function. This is to provide a means of transmitting payments. Thus by the use of debit cards, credit cards, Internet and telephone banking, cheques, direct debits, etc., money can be transferred from one person or institution to another without having to rely on cash.

Which of the above are examples of economies of scale?

Draw up a list of financial intermediaries and their functions in your local area.

Definitions

Financial intermediaries Financial institutions acting as a means of channelling funds from depositors to borrowers.

Maturity transformation The transformation of deposits into loans of a longer maturity.

Risk transformation The ability of financial institutions to spread risks by having a large number of clients.

and negotiate individual terms with the firm to suit the firm's particular requirements.

In the past, there were many independent wholesale banks, known as investment banks. These included famous names such as Morgan Stanley, Rothschild, S G Hambros and Goldman Sachs. With the worldwide financial crisis of 2008, however, most of the independent investment banks merged with universal banks, which conduct both retail and wholesale activities.

Functional separation of retail and wholesale banking. One particular concern about the rise of large universal banks is that the core activities of retail banks need

KI 15 p 80

isolating from risky wholesale banking activities. Otherwise, there is the danger that losses from wholesale activities may impact on important retail activities such as accepting deposits, facilities for withdrawing money or making payments from deposit accounts and the provision of overdraft facilities.

One way of achieving this is through the **functional separation** of retail and wholesale banking, such as by enforcing a ring fence between the different divisions of a financial institution. If losses are made by the wholesale division, these will not affect the retail division.

The ring-fencing of UK banking groups with more than £25 billion of core retail deposits became effective from January 2019. This followed the passing of the Financial Services (Banking Reform) Act in December 2013. Banks whose divisions are functionally separated from each other are known as 'ring-fenced banks' (RFBs).

Building societies

Building societies are UK institutions that historically have specialised in granting loans (mortgages) for house purchase. They compete for the savings of the general public through a network of high street branches. Unlike banks, they are not public limited companies, their 'shares' being the deposits made by their investors. In recent years, many of the building societies have converted to banks (including all the really large building societies except the Nationwide).

In the past, there was a clear distinction between banks and building societies. Today, however, they have become much more similar, with building societies now offering current account facilities and cash machines, and retail banks granting mortgages. As with the merging of retail and wholesale banks, this is all part of a trend away from the narrow specialisation of the past and towards the offering of a wider and wider range of services. This was helped by a process of **financial deregulation**.

MFIs

Banks and building societies are both examples of what are called **monetary financial institutions** (**MFIs**). This term is used to describe all deposit-taking institutions which, as we will see later, also includes countries' central banks (e.g. the Bank of England).

Deposit taking and lending

Balance sheets

Banks and building societies provide a range of **financial instruments**. These are financial claims, either by customers on the bank (e.g. deposits) or by the bank on its customers (e.g. loans). They are best understood by analysing the balance sheets of financial institutions, which itemise their liabilities and assets.

KI 32 p 239

A financial institution's liabilities are those financial instruments involving a financial claim on the financial institution itself. As we shall see, these are largely *deposits* by customers, such as current and savings accounts. Its assets are financial instruments involving a financial claim on a third party: these are *loans,* such as personal and business loans and mortgages.

The total liabilities and assets for UK MFIs are set out in the balance sheet in Table 11.1. The aggregate size of the balance sheet in 2022 was equivalent to around 4 times the UK's annual GDP. This is perhaps the simplest indicator of the significance of banks in modern economies, like the UK.

Both the *size* and *composition* of banks' balance sheets have become the focus of the international community's effort to ensure the stability of countries' financial systems. The growth of the aggregate balance sheet in the UK is considered in Box 11.1. We now focus on the composition of the balance sheet looking in more detail at the various types of liabilities and assets.

Liabilities

Customers' deposits in banks and building societies are **liabilities** to these institutions. This means simply that the customers have the claim on these deposits and thus the institutions are liable to meet the claims.

There are four major types of deposit: sight deposits, time deposits, certificates of deposit and 'repos'.

Sight deposits. **Sight deposits** are any deposits that can be withdrawn on demand by the depositor without penalty. In the past, sight accounts did not pay interest.

Definitions

Functional separation of banks The ring-fencing by banks of core retail banking services, such as deposit-taking, from riskier investment banking activities.

Financial deregulation The removal of or reduction in legal rules and regulations governing the activities of financial institutions.

Monetary financial institutions (MFIs) Deposit-taking financial institutions including banks, building societies and central banks.

Financial instruments Financial products resulting in a financial claim by one party over another.

Liabilities All legal claims for payment that outsiders have on an institution.

KI 32 p 249

Table 11.1 Balance sheet of UK MFIs (end of May 2022)

Sterling liabilities	£bn	%	Sterling assets	£bn	%
Sight deposits		52.8	Notes and coins	10.1	0.2
UK MFIs	112.6		Balances with UK central bank		20.2
UK public sector	20.3		Reserve balances	904.6	
UK private sector	2019.3		Cash ratio deposits	13.0	
Non-residents	225.7		Loans		7.5
Time deposits		24.9	UK MFIs	221.5	
UK MFIs	301.0		UK MFIs' CDs, etc.	2.7	
UK public sector	17.0		Non-residents	119.1	
UK private sector	643.0		Bills and acceptances	8.3	0.2
Non-residents	159.8		Reverse repos	386.6	8.5
Repos	200.4	4.5	Investments	431.1	9.5
CDs and other short-term papers	312.3	6.9	Advances	2376.6	52.2
Capital and other internal funds	410.4	9.1	Other assets	76.7	1.7
Other liabilities	79.8	1.8			
Total sterling liabilities	**4501.6**	100.0	**Total sterling assets**	**4550.3**	100.0
Foreign currency liabilities	4969.4		Total foreign currency assets	4920.6	
Total liabilities	9470.9		Total assets	9470.9	

Source: Based on data in *Bankstats* (Bank of England), Table B1.4, Data published 1 July 2022

Today, however, there are many sight accounts that do. In fact, there is quite aggressive competition nowadays between banks to offer apparently very attractive interest rates on such accounts, although these are often on balances up to a relatively small amount.

The most familiar form of sight deposits are current accounts at banks. Depositors are issued with cheque books and/or debit cards (e.g. Visa debit or MasterCard's Maestro) that enable them to spend the money directly without first having to go to the bank and draw the money out in cash. In the case of debit cards, the person's account is electronically debited when the purchase is made. This process is known as EFTPOS (electronic funds transfer at point of sale). Money can also be transferred between individuals and businesses through direct debits, standing orders and Internet banking transfers.

An important feature of current accounts is that banks often allow customers to be overdrawn. That is, they can draw on their account and make payments to other people in excess of the amount of money they have deposited.

Time deposits. **Time deposits** require notice of withdrawal. However, they normally pay a higher rate of interest than sight accounts. With some types of account, a depositor can withdraw a certain amount of money on demand, but there will be a penalty of so many days' lost interest. They are not cheque-book or debit-card accounts, although some allow customers to use cash cards. The most familiar form of time deposits are the deposit and savings accounts in banks and the various savings accounts in building societies. No overdraft facilities exist with time deposits.

A substantial proportion of time deposits are from the *banking sector*: i.e. other banks and other financial institutions. Interbank lending grew over the years as money markets were deregulated and as deposits increasingly moved from one currency to another to take advantage of different rates of interest between different countries. A large proportion of overseas deposits are from foreign banks.

Certificates of deposit. **Certificates of deposit (CDs)** are certificates issued by banks to customers (usually firms) for large deposits of a fixed term (e.g. £100 000 for

Definitions

Sight deposits Deposits that can be withdrawn on demand without penalty.

Time deposits Deposits that require notice of withdrawal or where a penalty is charged for withdrawals on demand.

Certificates of deposit (CDs) Certificates issued by banks for fixed-term interest-bearing deposits. They can be resold by the owner to another party.

18 months). They can be sold by one customer to another, and thus provide a means whereby the holders can get money quickly if they need it without the *banks* that have issued the CD having to supply the money. (This makes them relatively 'liquid' to the depositor but 'illiquid' to the bank: we examine this below.) The use of CDs has grown rapidly in recent years. Their use by firms has meant that, at a wholesale level, sight accounts have become *less* popular.

Sale and repurchase agreements (repos). If banks have a temporary shortage of funds, they can sell some of their financial assets to other banks or to the central bank – the Bank of England in the UK and the European Central Bank in the eurozone (see below) – and later repurchase them on some agreed date, typically a fortnight later. These **sale and repurchase agreements** (**repos**) are in effect a form of loan – the bank borrowing for a period of time using some of its financial assets as the security for the loan. One of the major assets to use in this way are government bonds, normally called 'gilt-edged securities' or simply 'gilts' (see below). Sale and repurchase agreements involving gilts are known as *gilt repos*. Gilt repos play a vital role in the operation of monetary policy (see Section 13.2).

Capital and other funds. This consists largely of the share capital in banks. Since shareholders cannot take their money out of banks, it provides a source of funding to meet sudden increases in withdrawals from depositors and to cover bad debts.

It is vital that banks have sufficient capital. As we shall see, an important part of the response to the financial crisis has been to require banks to hold relatively larger amounts of capital. In 2022, the aggregate amount of sterling capital held by banks based in the UK was equivalent to around 10 per cent of their sterling liabilities (see Table 11.1).

Assets

A bank's financial **assets** are its claims on others. There are three main categories of assets.

Cash and reserve balances in the central bank (Bank of England in the UK, ECB in the eurozone). Banks need to hold a certain amount of their assets as cash. This is largely used to meet the day-to-day demands of customers.

They also keep 'reserve balances' in the central bank. In the UK these earn interest at the Bank of England's repo rate (or 'Bank Rate' as it is called), if kept within an agreed target range. These are like the banks' own current accounts and are used for clearing purposes (i.e. for settling the day-to-day payments between banks). They can be withdrawn in cash on demand. With interbank lending being seen as too risky during the crisis of 2008, many banks resorted to depositing surplus cash in the Bank of England, even though the Bank Rate was lower than interbank interest rates.

In the UK, banks and building societies are also required to deposit a small fraction of their assets as 'cash ratio deposits' (CRDs) with the Bank of England. These cannot be drawn on demand and earn no interest.

As you can see from Table 11.1, cash and balances in the Bank of England account for around 20 per cent of banks' assets. This percentage has grown in recent years as the Bank of England has created additional money through a process known as quantitative easing (see Box 13.6 on page 396). The majority of banks' assets are in the form of various types of loan. These are 'assets' because they represent claims that the banks have on other people.

Loans can be grouped into two types: short and long term.

Short-term loans. These are in the form of *market loans*, *bills of exchange* or *reverse repos*. The market for these various types of loan is known as the **money market**.

- **Market loans** are made primarily to other banks or financial institutions. This interbank lending consists of (a) money lent 'at call' (i.e. reclaimable on demand or at 24 hours' notice); (b) money lent for periods up to one year, but typically a few weeks; (c) CDs (i.e. certificates of deposits made in other banks or building societies).
- **Bills of exchange** are loans either to companies (**commercial bills**) or to the government (**Treasury bills**). These are, in effect, an IOU, with the company issuing them (in the case of commercial bills), or the Bank of England on behalf of the government (in the case of Treasury bills), promising to pay the holder a specified sum on a particular date (the 'maturity date'), typically three months later.

Definitions

Sale and repurchase agreements (repos) An agreement between two financial institutions whereby one in effect borrows from another by selling its assets, agreeing to buy them back (repurchase them) at a fixed price and on a fixed date.

Assets Possessions or claims held on others.

Money market The market for short-term loans and deposits.

Market loans Short-term loans (e.g. money at call and short notice).

Bills of exchange Certificates promising to repay a stated amount on a certain date, typically three months from the issue of the bill. Bills pay no interest as such, but are sold at a discount and redeemed at face value, thereby earning a rate of discount for the purchaser.

Commercial bills Bills of exchange issued by firms.

Treasury bills Bills of exchange issued by the Bank of England on behalf of the government. They are a means whereby the government raises short-term finance.

BOX 11.2 THE GROWTH OF BANKS' BALANCE SHEETS

The rise of wholesale funding

Banks' traditional funding model relied heavily on deposits as the source of funds for loans. However, new ways for financial institutions to access funds to generate new loans evolved. These reflected the deregulation of financial markets and the rapid pace of financial innovation.

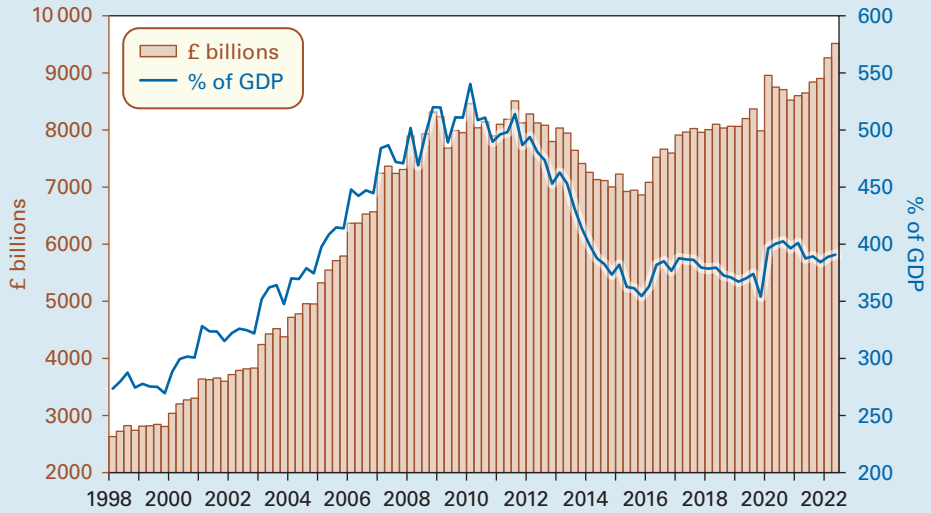
Seeds of the 2007–8 crisis

Increasingly financial institutions made greater use of *wholesale funds*. These are funds obtained mainly from other financial institutions. This coincided with the growth in a process known as *securitisation*. This involves the conversion of non-marketable banks' assets, such as residential mortgages, which have regular income streams (e.g. from payments of interest and capital), into assets that could be traded, known as 'tradable financial instruments'. These provide lenders who originate the loans with a source of funds for further loans. Therefore, securitisation became another means of obtaining funds from other financial institutions. Securitisation is discussed further in Box 11.3.

With an increasing use of money markets by financial institutions, vast sums of funds became available for lending. One consequence of this is illustrated in the chart: the expansion of the aggregate balance sheet. The balance sheet grew from £2.6 trillion ($2\frac{3}{4}$ times GDP) in 1998 to £8.5 trillion ($5\frac{1}{2}$ times GDP) in 2010.

KI 32 p 239

Aggregate balance sheet of UK banks and building societies



Note: Since 2010 all loans securitised by MFIs are recorded on MFI balance sheets

Sources: (i) Data showing liabilities of banks and building societies based on series LPMALOA and RPMTBJF (up to the end of 2009) and RPMB3UQ (from 2010) from *Statistical Interactive Database*, Bank of England (data published 30 September 2022, not seasonally adjusted). (ii) GDP data based on series YBHA, Office for National Statistics (GDP figures are the sum of the latest four quarters)

Since bills do not pay interest, they are sold below their face value (at a 'discount') but redeemed on maturity at the face value. This enables the purchaser, in this case the bank, to earn a return. The market for new or existing bills is therefore known as the **discount market**.

The price paid for bills will depend on demand and supply. For example, the more Treasury bills that are offered for sale (i.e. the higher the supply), the lower their equilibrium price will be, and hence the higher their rate of return (i.e. their rate of interest, or 'rate of discount') will be.

- **Reverse repos.** When a sale and repurchase agreement is made, the financial institution *purchasing* the assets (e.g. gilts) is, in effect, giving a short-term loan. The other party agrees to buy back the assets (i.e. pay back the loan) on a set date. The assets temporarily held by the bank making the loan are known as 'reverse repos'. Reverse repos are typically

Definitions

Discount market An example of a money market in which new or existing bills are bought and sold.

Reverse repos Gilts or other assets that are purchased under a sale and repurchase agreement. They become an asset to the purchaser.

EXPLORING ECONOMICS

The growth in banks' balance sheets was accompanied by a change in their composition.

First, the profile of banks' assets became less liquid as they extended more long-term credit to households and firms. Assets generally became riskier, as banks increasingly granted mortgages of 100 per cent or more of the value of houses – a problem for banks if house prices fell and they were forced to repossess.

Second, there was a general increase in the use of fixed-interest bonds as opposed to ordinary shares (equities) for raising capital. The ratio of bonds to equity capital is known as **gearing (or leverage) ratio**. The increase in leverage meant that banks were operating with lower and lower levels of loss-absorbing capital, such as ordinary shares. If banks run at a loss, dividends on shares can be suspended; the payment of interest on fixed interest bonds cannot. This meant that as the crisis unfolded, policy makers were facing a liquidity problem, not among one or two financial institutions, but across the financial system.

KI 22 p 153

The market failure we are describing is a form of **co-ordination failure** and is an example of the *fallacy of composition*. When one bank pursues increased earnings by borrowing from and lending to other financial institutions, this is not necessarily a problem. But, if many institutions build their balance sheets by borrowing from and lending to *each other,* then it becomes a problem for the whole financial system. The apparent increase in liquidity for individual banks, on which they base credit, is not an overall increase in liquidity for the financial system as a whole. The effect is to create a credit bubble.

KI 32 p 239

The dangers of the bubble for the financial system and beyond were magnified by the increasingly tangled web of interdependencies between financial institutions, both nationally and globally. There was a danger that this complexity was masking fundamental weaknesses of many financial institutions and too little overall liquidity.

The financial crisis

Things came to a head in 2007 and 2008. Once one or two financial institutions failed, such as Northern Rock in the UK in September 2007 and Lehman Brothers in the USA in September 2008, the worry was that failures would spread like a contagion. Banks could no longer rely on each other as their main source of liquidity.

The problems arising from the balance sheet expansion, increased leverage and a heightened level of maturity mismatch meant that central banks around the world, including the Bank of England, were faced with addressing a liquidity problem of huge proportions. They had to step in to supply central bank money to prevent a collapse of the banking system.

Subsequently, the Basel Committee on Banking Supervision (see pages 313–17) agreed a set of measures, to be applied globally, designed to ensure the greater financial resilience of banks and banking systems.

The chart shows that from the 2010s there was a decline in the size of the aggregate balance sheet of banks resident in the UK from around 5½ times to between 3½ and 4 times GDP.

Why do you think banks became reluctant to deposit money with other banks during the financial crisis of the late 2000s?

Using the **Bank of England Statistical Interactive Database** *download monthly data on MFI total sterling assets* **(RPMB3XP)**. *Construct a bar chart showing the series across time. Briefly summarise your findings.*

Definitions

Gearing or **leverage** (US term) The ratio of debt capital to equity capital: in other words, the ratio of borrowed capital (e.g. bonds) to shares.

Co-ordination failure When a group of firms (e.g. banks) acting independently could have achieved a more desirable outcome if they had co-ordinated their decision making.

for one week but can be for as little as overnight to as long as one year.

Longer-term loans. These consist primarily of loans to customers, both personal customers and businesses. These loans, also known as *advances*, are of four main types: fixed-term (repayable in instalments over a set number of years – typically, six months to five years), overdrafts (often for an unspecified term), outstanding balances on credit-card accounts and mortgages (typically for 25 years).

Investments. Banks also make investments. These are partly in government bonds (gilts), which are effectively loans to the government. The government sells bonds, which then pay a fixed sum each year as interest. Once issued, they can then be bought and sold on the stock exchange. Banks are normally only prepared to buy bonds that have less than five years to maturity (the date when the government redeems the bonds). Banks also invest in other financial institutions, including subsidiary financial institutions.

Taxing the banks

Bank levy. In January 2011, the UK introduced the bank levy: a tax on the *liabilities* of banks and building societies operating in the UK. The tax was to be applied to the global balance sheet of banks with their HQs in the UK and for subsidiaries of non-UK banks just to their UK activities.

KI 9 p 26

The levy is founded on two key principles. First, the revenues raised should be able to meet the full fiscal costs of any future support for financial institutions. Second, it should provide banks with incentives to reduce risk-taking behaviour and so reduce the likelihood of future financial crises.

KI 7 p 24

The UK bank levy has two rates: a full rate on taxable liabilities with a maturity of less than 1 year and a half rate on taxable liabilities with a maturity of more than 1 year. This differential is intended to discourage excessive short-term borrowing by the banks in their use of wholesale funding.

Not all liabilities are subject to the levy. First, it is not imposed on the first £20 billion of liabilities. This is to encourage small banks (note that the largest UK banks, such as HSBC, Barclays and RBS, each have liabilities of over £2 trillion). Second, various liabilities are excluded. These are: (a) gilt repos; (b) retail deposits insured by public schemes such as the UK's Financial Services Compensation Scheme, which guarantees customers' deposits of up to £85 000; (c) a large part of a bank's capital known as Tier 1 capital (see below) – the argument here is that it is important for banks to maintain sufficient funds to meet the demands of its depositors.

Banks are also able to offset against their taxable liabilities holdings of highly liquid assets, such as Treasury bills and cash reserves at the Bank of England. These exclusions and deductions are designed to encourage banks to engage in less risky lending.

The levy rates were initially set at 0.075 and 0.0375 per cent, with the intention that the levy raised at least £2.5 billion each year. As the aggregate balance sheets of banks began to shrink, the rates were increased several times. Then, in the 2015 Budget it was announced that the levy rates, which at that time were 0.21 and 0.105 per cent, were to be gradually reduced until from 2021 they would be 0.1 and 0.05 per cent. Also, with concerns about the effect of the levy on the international competitiveness of UK global banks, it was announced that from 2021 the levy would apply only to UK balance sheets.

Bank corporation tax surcharge. The 2015 Budget also saw the announcement of a new 8 per cent corporation tax surcharge on banks. This marked a shift in the tax base from banks' balance sheets to their profits. The 8 per cent surcharge was introduced in January 2016 for all banks with annual profits over £25 million. However, with the main rate of corporation tax rising in April 2023 from 19 to 25 per cent (which was confirmed in the Autumn Statement of 2022 after earlier plans to abandon it), the government cut the surcharge to 3 per cent and raised the profit threshold to £100 million. This was again designed to address concerns that UK-headquartered banks would otherwise be put at a disadvantage.

Profitability, liquidity and capital adequacy

As we have seen, banks keep a range of liabilities and assets. The balance of items in this range is influenced by three important considerations: profitability, liquidity and capital adequacy.

Profitability

Profits are made by lending money out at a higher rate of interest than that paid to depositors. The average interest rate received by banks on their assets is greater than that paid by them on their liabilities.

Liquidity

The **liquidity** of an asset is the ease with which it can be converted into cash without loss. Cash itself, by definition, is perfectly liquid.

Some assets, such as money lent at call to other financial institutions, are highly liquid. Although not actually cash, these assets can be converted into cash virtually on demand with no financial penalty. Other short-term interbank lending is also very liquid. The only issue here is one of confidence that the money will actually be repaid. This was a worry in the financial crisis of 2008/9, when many banks stopped lending to each other on the interbank market for fear that the borrowing bank might become insolvent.

Other assets, such as gilts, can be converted into cash straight away by selling them on the Stock Exchange, but with the possibility of some financial loss, given that their market price fluctuates. Such assets, therefore, are not as liquid as money at call.

Other assets are much less liquid. Personal loans to the general public or mortgages can be redeemed by the bank only as each instalment is paid. This was why securitisation of mortgages became popular with banks as it effectively made their mortgage assets tradable and hence more liquid (see Box 11.3).

Banks must always be able to meet the demands of their customers for withdrawals of money. To do this, they must hold sufficient cash or other assets that can be readily turned into cash. In other words, banks must maintain sufficient liquidity.

The balance between profitability and liquidity

Profitability is the major aim of banks and most other financial institutions. However, the aims of profitability and liquidity tend to conflict. In general, the more liquid an asset, the less profitable it is, and *vice versa*.

Definition

Liquidity The ease with which an asset can be converted into cash without loss.

Personal and business loans to customers are profitable to banks, but highly illiquid. Cash is totally liquid, but earns no profit. Thus financial institutions like to hold a range of assets with varying degrees of liquidity and profitability.

For reasons of *profitability,* banks will want to 'borrow short' (at low rates of interest, such as on current accounts) and 'lend long' (at higher rates of interest, such as on personal loans). The difference in the average maturity of loans and deposits is known as the **maturity gap**. In general terms, the larger the maturity gap between loans and deposits, the greater the profitability. For reasons of *liquidity,* however, banks will want a relatively small gap: if there is a sudden withdrawal of deposits, banks will need to be able to call in enough loans.

The ratio of an institution's liquid assets to total assets (or liabilities) is known as its **liquidity ratio.** For example, if a bank had £100 million of assets, of which £10 million were liquid and £90 million were illiquid, the bank would have a 10 per cent liquidity ratio.

If a financial institution's liquidity ratio is too high, it will make too little profit. If the ratio is too low, there will be the risk that customers' demands may not be able to be met: this would cause a crisis of confidence and possible closure. Institutions thus have to make a judgement as to what liquidity ratio is best – one that is neither too high nor too low.

Balances in the central bank, short-term loans (i.e. those listed above) and government bonds with less than 12 months to maturity (and hence tradable now at near to their face value) would normally be regarded as liquid assets.

KI 15
p 80

Over the years, banks had reduced their liquidity ratios (i.e. the ratio of liquid assets to total assets). This was not a problem as long as banks could always finance lending to customers by borrowing on the interbank market. In the late 2000s, however, banks became increasingly worried about bad debt. They thus felt the need to increase their liquidity ratios and hence cut back on lending and chose to keep a higher proportion of deposits in liquid form. In the UK, for example, banks substantially increased their level of reserves in the Bank of England.

Pause for thought

Why are government bonds that still have 11 months to run regarded as liquid, whereas overdrafts granted for a few weeks are not?

Secondary marketing and securitisation

As we have seen, one way of reconciling the two conflicting aims of liquidity and profitability is for financial institutions to hold a mixture of liquid and illiquid assets. Another way is through the **secondary marketing** of assets. This is where holders of assets sell them to someone else before the maturity date. This allows banks to close the maturity gap for *liquidity* purposes, but maintain the gap for *profitability* purposes.

Certificates of deposit (CDs) are a good example of secondary marketing. CDs are issued for fixed-period deposits in a bank (e.g. one year) at an agreed interest rate. The bank does not have to repay the deposit until the year is up. CDs are thus illiquid liabilities for the bank, and they allow it to increase the proportion of illiquid assets without having a dangerously high maturity gap. But the holder of the CD in the meantime can sell it to someone else (through a broker). It is thus liquid to the holder. Because CDs are liquid to the holder, they can be issued at a relatively *low* rate of interest and thus allow the bank to increase its profitability.

Another example of secondary marketing is when a financial institution sells some of its assets to another financial institution. The advantage to the first institution is that it gains liquidity. The advantage to the second one is that it gains profitable assets. The most common method for the sale of assets has been through a process known as **securitisation**.

Securitisation occurs when a financial institution pools some of its assets, such as residential mortgages, and sells them to an intermediary known as a **Special Purpose Vehicle** (**SPV**). SPVs are legal entities created by the financial institution. In turn, the SPV funds its purchase of the assets by issuing bonds to investors (noteholders). These bonds are known as **collateralised debt obligations** (**CDOs**). The sellers (e.g. banks) get cash now rather than having to wait and can use it to fund loans to customers. The buyers make a profit if the income yielded by the CDOs are as expected. Such bonds can be very risky, however, as the future cash flows may be *less* than anticipated.

Definitions

Maturity gap The difference in the average maturity of loans and deposits.

Liquidity ratio The proportion of a bank's total assets held in liquid form.

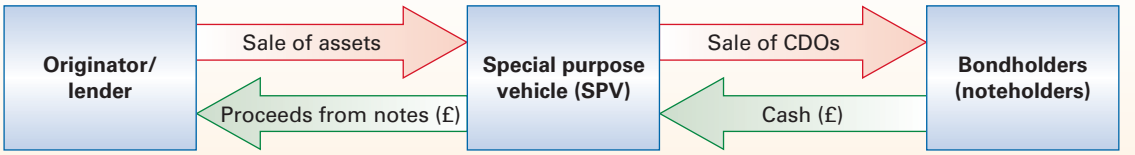
Secondary marketing Where assets are sold before maturity to another institution or individual.

Securitisation Where future cash flows (e.g. from interest rate or mortgage payments) are turned into marketable securities, such as bonds.

Special purpose vehicle (**SPV**) Legal entities created by financial institutions for conducting specific financial functions, such as bundling assets together into fixed interest bonds and selling them.

Collateralised debt obligations (**CDOs**) These are a type of security consisting of a bundle of fixed-income assets, such as corporate bonds, mortgage debt and credit-card debt.

Figure 11.1 Securitisation chain



The securitisation chain is illustrated in Figure 11.1. The financial institution looking to sell its assets is referred to as the 'originator' or the 'originator-lender'. Working from left to right, we see the originator-lender sells its assets to another financial institution, the SPV, which then bundles assets together into CDOs and sells them to investors (e.g. banks or pension funds) as bonds. Now working from right to left, we see that by purchasing the bonds issued by the SPV, the investors provide the funds for the SPV's purchase of the lender's assets. The SPV is then able to use the proceeds from the bond sales (CDO proceeds) to provide the originator-lender with liquidity.

The effect of secondary marketing is to reduce the liquidity ratio that banks feel they need to keep. It has the effect of increasing their maturity gap.

Dangers of secondary marketing. There are dangers to the banking system, however, from secondary marketing. To the extent that banks individually feel that they can operate with a lower liquidity ratio, so this will lead to a lower national liquidity ratio. This may lead to an excessive expansion of credit (illiquid assets) in times of economic boom.

KI 15 p 80

Also, there is an increased danger of banking collapse. If one bank fails, this will have a knock-on effect on those banks which have purchased its assets. In the case of securitisation, the strength of the chain is potentially weakened if individual financial institutions move into riskier market segments, such as **sub-prime** residential mortgage markets. Should the income streams of the originator's assets dry-up – for instance, if individuals default on their loans – then the impact is felt by the whole of the chain. In other words, institutions and investors are exposed to the risks of the originator's lending strategy.

The issue of securitisation and its impact on the liquidity of the financial system during the 2000s is considered in Box 11.3.

Capital adequacy

KI 15 p 80

In addition to sufficient liquidity, banks must have sufficient capital (i.e. funds) to allow them to meet all demands from depositors and to cover losses if borrowers default on payment. Capital adequacy is a measure of a bank's capital relative to its assets, where the assets are weighted according to the degree of risk. The riskier the assets, the greater the amount of capital that will be required.

A measure of capital adequacy is given by the **capital adequacy ratio** (**CAR**). This is the ratio of suitable capital to risk-weighted assets:

$$\text{CAR} = \frac{\text{Common Equity Tier 1 capital} + \text{Additional Tier 1 capital} + \text{Tier 2 capital}}{\text{Risk-weighted assets}}$$

Common Equity Tier 1 (CET1) capital includes bank reserves (from retained profits) and ordinary share capital (equities), where dividends to shareholders vary with the amount of profit the bank makes. Such capital thus places no burden on banks in times of losses as no dividend need be paid. What is more, unlike depositors, shareholders cannot ask for their money back.

Additional Tier 1 (AT1) capital consists largely of preference shares. These pay a fixed dividend (like company bonds), but although preference shareholders have a prior claim over ordinary shareholders on company profits, dividends need not be paid in times of loss.

Tier 2 capital is 'subordinated debt' with a maturity greater than 5 years. Subordinated debt holders only have a claim on a company after the claims of all other bondholders have been met.

Risk-weighted assets are the total value of assets, where each type of asset is multiplied by a risk factor. Under the internationally agreed Basel II accord, cash and government bonds have a risk factor of zero and are thus not included. Interbank lending between the

Definitions

Sub-prime debt Debt where there is a high risk of default by the borrower (e.g. mortgage holders who are on low incomes facing higher interest rates and falling house prices).

Capital adequacy ratio (**CAR**) The ratio of a bank's capital (reserves and shares) to its risk-weighted assets.

major banks has a risk factor of 0.2 and is thus included at only 20 per cent of its value; residential mortgages have a risk factor of 0.35; personal loans, credit-card debt and overdrafts have a risk factor of 1; loans to companies carry a risk factor of 0.2, 0.5, 1 or 1.5, depending on the credit rating of the company. Thus the greater the average risk factor of a bank's assets, the greater will be the value of its risk-weighted assets, and the lower will be its CAR.

The greater the CAR, the greater the capital adequacy of a bank. Under Basel II, banks were required to have a CAR of at least 8 per cent (i.e. 0.08). They were also required to meet two supplementary CARs. Firstly, banks needed to hold a ratio of Tier 1 capital to risk-weighted assets of at least 4 per cent and, secondly, a ratio of ordinary share capital to risk-weighted assets of at least 2 per cent. It was felt that these three ratios would provide banks with sufficient capital to meet the demands from depositors and to cover losses if borrowers defaulted. The financial crisis, however, meant a rethink.

Strengthening international regulation of capital adequacy and liquidity

KI 15 p 80

In the light of the financial crisis of 2007–9, international capital adequacy requirements were strengthened by the *Basel Committee on Banking Supervision* with a final agreement in December 2017. The 'Basel III' capital requirements, as they are called, were phased in, with the new minimum capital adequacy ratios operating from 2019. They are summarised in Figure 11.2.

Raising minimum levels of capital adequacy

The new framework raises the minimum levels of capital adequacy of all financial institutions. From its introduction in 2013, banks continued to need a CAR of at least 8 per cent (i.e. 0.08). However, from 2015 they were required to operate with a ratio of CET1 to risk-weighted assets of at least 4.5 per cent. Then from 2016 began a phased introduction of a *capital conservation buffer* raising the CET1 ratio to no less than 7 per cent from 2019. This increases the minimum CAR therefore to at least 10.5 per cent. If a bank breaches the capital conservation buffer, limitations are placed on dividend and bonus payments.

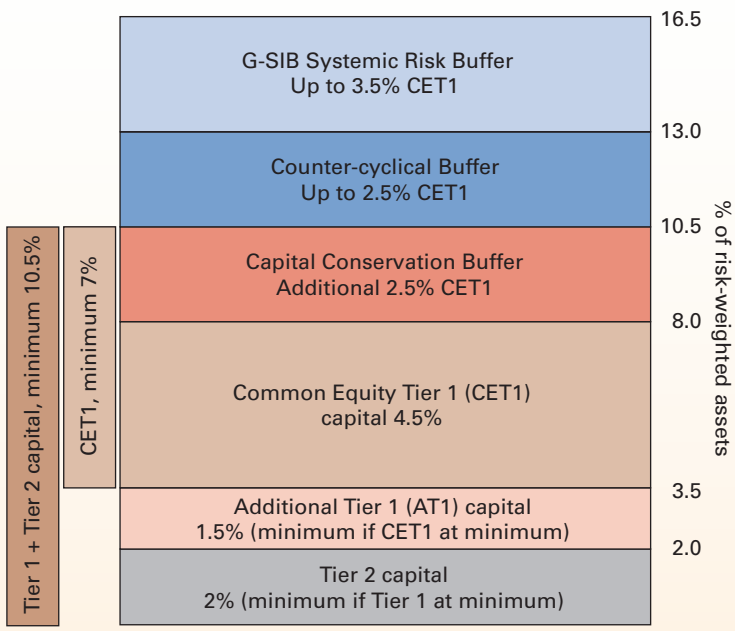
Counter-cyclical buffer

The new framework places more emphasis on national regulators assessing the financial resilience across all financial institutions under its jurisdiction, particularly in the context of the macroeconomic environment. This is the principle of **macro-prudential regulation**.

Definition

Macro-prudential regulation Regulation which focuses on the financial system as a whole and which monitors its impact on the wider economy.

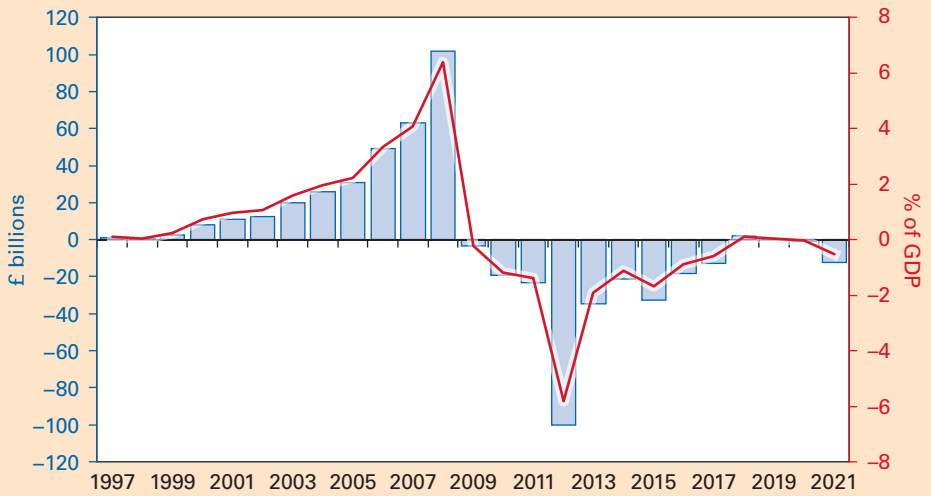
Figure 11.2 Basel III minimum capital requirements



BOX 11.3 THE RISE OF SECURITISATION

Spreading the risk or promoting a crisis?

Net securitisations of secured lending to individuals



Note: Data up to 2010 relate to changes in other specialist lenders' sterling net securitisations of secured lending to individuals and housing associations. From 2010 data relate to changes in resident MFI sterling securitised loans secured on dwellings to individuals

Sources: Based on data from *Statistical Interactive Database*, Bank of England, series LPMVUJD (up to 2010) and LPMB8GO (data published 29 March 2022) and series YBHA, National Statistics

The conflict between profitability and liquidity may have sown the seeds for the credit crunch that affected economies across the globe in the second half of the 2000s.

To understand this, consider the size of the 'advances' item in the UK banking sector's balance sheet – just over half of the value of sterling assets (see Table 11.1). The vast majority of these are to households. Advances secured against property have, in recent times, accounted for around 80 per cent by value of credit advanced by financial institutions to households. *Residential mortgages* involve institutions lending long.

Securitisation of debt

One way in which individual institutions can achieve the necessary liquidity to expand the size of their mortgage lending (illiquid assets) is through *securitisation*. Securitisation grew especially rapidly in the UK and USA. In the UK this was particularly true among banks; building societies have historically made greater use of retail deposits to fund advances.

Securitisation is a form of financial engineering. It provides banks (originator-lenders) with liquidity and enables them to engage in further lending opportunities. It provides the Special Purpose Vehicles (SPVs) with the opportunity to issue profitable securities.

In the period up to 2010 most securitisations in the UK saw the original loans moving off the balance sheet of MFIs and onto the balance sheet of the SPV issuing the Collateralised Debt Obligations (CDOs). From 2010, however, all securitisations are detailed on the balance sheets of

If necessary, regulators can apply a *counter-cyclical buffer* (CCyB) to all banks, so increasing the CET1 ratio by up to a further 2.5 per cent. This allows financial institutions to build up a capital buffer during periods of high economic growth to allow it to be drawn on in times of recession or financial difficulty. It should help to prevent financial institutions destabilising the economy by generating excessively strong credit cycles (see Section 9.6). In March 2020, for example, the Bank of England's Financial Policy Committee reduced the CCyB from 1 to 0 per cent to encourage banks to lend, thereby absorbing rather than amplifying the impact of the COVID shock. As inflation rates rose in 2022 and the macroeconomic environment deteriorated, the Bank raised the buffer to 1 per cent from December 2022 and then to 2 per cent from July 2023.

Pause for thought

Could a countercyclical buffer fall during periods with a positive output gap?

Capital buffers for systemically important banks

Global systemically important banks. Large global financial institutions, known as **global systemically important banks** (**G-SIBs**), are required to operate with a CET1 ratio of up to 3.5 per cent

CASE STUDIES & APPLICATIONS

monetary financial institutions (MFIs), including previous securitisations which, as a result, have been brought back on to the balance sheets of MFIs.

The chart shows the rapid growth in the flows of securitised *secured* loans in the UK from £1 billion in 1997 to over £100 billion by 2008: the equivalent of an increase from 0.1 to 6.5 per cent of GDP. This increase reflected the strong demand among investors for CDOs. The attraction of these fixed-income products for the noteholders was the potential for higher returns than on (what were) similarly rated products. However, investors have no recourse should people with mortgages fall into arrears or, worse still, default on their mortgages.

KI 15 p 80

Risks and the sub-prime market

The securitisation of assets is not without risks for all those in the securitisation chain and consequently for the financial system as a whole.

KI 15 p 80

The pooling of advances in itself *reduces* the cash-flow risk facing investors. However, there is a **moral hazard** problem here (see also page 83). The pooling of the risks may encourage originator-lenders to lower their credit criteria by offering higher income multiples (advances relative to annual household incomes) or higher loan-to-value ratios (advances relative to the price of housing).

Towards the end of 2006 the USA witnessed an increase in the number of defaults by households on residential mortgages. This was a particular problem in the sub-prime market – higher-risk households with poor credit ratings. Similarly, the number falling behind with their payments rose. This was on the back of rising interest rates.

These problems in the US sub-prime market were the catalyst for the liquidity problem that beset financial systems in 2007 and 2008. Where these assets were securitised, investors, largely other financial institutions, suffered from the contagion arising from arrears and defaults.

Securitisation also spread the contagion across the world. Investors are global, so that supporting CDOs for, say, an American family's residential mortgage, are effectively travelling across national borders. This resulted in institutions writing-off debts, a deterioration of their balance sheets, the collapse in the demand for securitised assets and the drying-up of liquidity.

The chart shows that during the 2010s banks were *buying back* CDOs from SPVs (negative net securitisations). A similar pattern was observed in many other countries. This contributed to an easing in the growth of banks' balance sheets (see Box 11.2). This process was accelerated by central banks, which began to accept securitised assets in exchange for liquidity, either as part of programmes of quantitative easing to increase the amount of money in the economy (see Box 13.6) or other liquidity insurance mechanisms (see pages 317–18).

Does securitisation necessarily involve a moral hazard problem?

Using the Bank of England Statistical Interactive Database download monthly data on the stocks of securitised loans secured on dwellings (LPMB7GT) and of other securitised loans (LPMB7GU). Construct a stacked column chart showing the size and composition of the securitised debt stock of MFIs resident in the UK.

Definition

Moral hazard The temptation to take more risks when you know that someone else will cover the risks if you get into difficulties. In the case of banks taking risks, the 'someone else' may be another bank, the central bank or the government.

higher than other banks, depending on banks' size and global reach. This additional capital buffer is known as the *G-SIB systemic risk buffer*. The buffer recognises the likelihood that the failure of such larger institutions could trigger a global financial crisis.

With a G-SIB systemic risk buffer of potentially up to 3.5 per cent, the overall CAR for very large financial institutions can be as high as 16.5 per cent (see Figure 11.2).

Domestic systemically important banks. Several countries have also adopted a systemic risk buffer on financial institutions identified by their national authorities as **domestic systemically important banks** (D-SIBs).

Definitions

Global systemically important banks (G-SIBs) Banks identified by a series of indicators as being significant banks in the global financial system.

Domestically systemically important banks (D-SIBs) Banks identified by national regulators as being significant banks in the domestic financial system.

In the UK it was decided that individual ring-fenced banks (RFBs) and large building societies which hold more than £25 billion in deposits would be required to hold an additional systemic risk buffer. UK bank groups identified as G-SIBs containing a ring-fenced bank are therefore subject to two systemic risk buffers: the systemic risk buffer for G-SIBs at the group level and the domestic systemic risk buffer at the level of the ring-fenced bank.

The UK's domestic systemic risk buffer depends on the size of risk-weighted assets, with the initial plan for additional capital requirement of up to 3 per cent of these assets. However, depending on the judgment of the Financial Policy Committee, whose job is to ensure the financial resilience of the UK financial sector, the asset bands and the systemic risk buffer rates applied can be reviewed and adjusted if necessary.

Leverage ratio

To supplement the risk-based capital requirements, the Basel III framework introduced from 2018 a *non-risk-based leverage ratio* (i.e. separate from the ratios listed in Figure 11.2). This was described by regulators as complementary to the risk-based framework, it is intended to prevent a build-up of debt to fund banks' activities.

The leverage ratio buffer requires financial institutions to operate with a Tier 1 capital-to-asset ratio of 3 per cent. In contrast to the risk-based ratios, the assets in the denominator in this ratio are *not* weighted by risk factors.

Standardising the measure of risk

The impact of the Basel III's capital adequacy requirements was to increase the capital cushions of many banks significantly. Indeed, it seemed that banks were generally making good progress, with some meeting the requirements well ahead of time.

Despite this, an issue that was to remain outstanding for some time was the *calculation* of risk-weighted assets. Because of the complexity of banks' asset structures, which tend to vary significantly from country to country, it can be difficult to ensure that banks are meeting the Basel III requirements. The problem can then be made worse by the use of internal models by financial institutions which can apply different approaches to assessing and evaluating risks.

It was therefore proposed that banks would have to compare their own calculations with a 'standardised model'. Their own calculations of risk-based assets would then not be allowed to fall below a set percentage, known as 'the output floor', of the standardised approach. The argument then centred on how high the output floor should be set.

Those arguing for a higher output floor pointed to the importance of a credible regulatory framework. An over-reliance on inconsistently applied internal risk assessment methods could undermine the credibility of the regulatory framework. However, others argued that a high output floor could penalise financial institutions by treating assets as equally risky across countries. For example, Germany argued that since mortgage defaults have been rare, German mortgage debt should be given a lower weighting than US mortgage debt, where defaults have been more common. Hence, by setting the output floor too high banks could be judged to be undercapitalised.

In December 2017 the Basel Committee finally agreed that the output floor for the risk weighting of assets would be set at 72.5 per cent, meaning that individual countries could not set the risk of an asset, such as a mortgage, at less than 72.5 per cent of the level set by international regulators in their standardised models. This will, however, be phased in with an output floor of 50 per cent from 2023 increasing each year until reaching 72.5 per cent in 2028.

Liquidity

The financial crisis drew attention to the need for banks not only to hold adequate levels of capital but also to manage their liquidity better. The Basel III framework includes a *liquidity coverage ratio* (LCR). This requires that financial institutions have high quality liquid assets (HQLAs) to cover the expected net cash flow over the next 30 days. From 2019, the minimum LCR ratio (HQLAs relative to the expected 30-day net cash flow) became 100 per cent.

Net stable funding ratio (NSFR)

As part of Basel III, a minimum *net stable funding ratio* (NSFR) became a regulatory standard from 2018. This takes a longer-term view of the funding profile of banks by focusing on the reliability of liabilities as a source of funds, particularly in circumstances of extreme stress. The NSFR is the ratio of stable liabilities to assets likely to require funding (i.e. assets where there is a likelihood of default).

On the liabilities side, these are weighted by their expected reliability – in other words, by the stability of these funds. This weighting reflects the maturity of the liabilities and the likelihood of lenders withdrawing their funds. For example, Tier 1 and 2 capital have a weighting of 100 per cent; term deposits with less than one year to maturity have a weighting of 50 per cent; and unsecured wholesale funding has a weighting of 0 per cent. The result of these weightings is a measure of stable funding.

KI 15 p 80

On the assets side, these are weighted by the likelihood that they will have to be funded over the course of one year. This means that they are weighted by

their liquidity, with more liquid assets requiring less funding. Thus cash has a zero weighting, while more risky assets have weightings of up to 100 per cent. The result is a measure of required funding.

Banks must hold a minimum stable-liabilities-to-required-funding ratio (NSFR) of 100 per cent.

The central bank

The Bank of England is the UK's central bank. The European Central Bank (ECB) is the central bank for the countries using the euro. The Federal Reserve System (the Fed) is the USA's central bank. All countries with their own currency have a central bank and they fulfil two vital roles in the economy.

The first is to oversee the whole monetary system and ensure that banks and other financial institutions operate as stably and as efficiently as possible.

The second is to act as the government's agent, both as its banker and in carrying out monetary policy.

The Bank of England traditionally worked in very close liaison with the Treasury, and there used to be regular meetings between the Governor of the Bank of England and the Chancellor of the Exchequer. Although the Bank may have disagreed with Treasury policy, it always carried it out. In 1997, however, the Bank of England was given independence to decide the course of monetary policy. In particular, this meant that the Bank of England and not the government would now decide interest rates.

Another example of an independent central bank is the European Central Bank, which operates the monetary policy for the eurozone countries. Similarly, the Fed is independent of both the President and Congress, and its chairman is generally regarded as having great power in determining the country's economic policy. Although the degree of independence of central banks from government varies considerably around the world, there has been a general move in recent years to make central banks more independent.

Within their two broad roles, central banks typically have a number of different functions. Although we will consider the case of the Bank of England, the same principles apply to other central banks, such as the ECB and the Fed.

It issues notes

The Bank of England is the sole issuer of banknotes in England and Wales (in Scotland and Northern Ireland, retail banks issue banknotes). The amount of banknotes issued by the Bank of England depends largely on the demand for notes from the general public. If people draw more cash from their bank accounts, the banks will have to draw more cash from their balances in the Bank of England.

It acts as a bank

To the government. It keeps the two major government accounts: the 'Exchequer' and the 'National Loans Fund'. Taxation and government spending pass through the Exchequer. Government borrowing and lending pass through the National Loans Fund. The government tends to keep its deposits in the Bank of England to a minimum. If the deposits begin to build up (from taxation), the government will probably spend them on paying back government debt. If, on the other hand, it runs short of money, it will simply borrow more.

To banks. Banks' deposits in the Bank of England consist of reserve balances and cash ratio deposits (see Table 11.1). The reserve balances are used for clearing purposes between the banks, but are also a means by which banks can manage their liquidity risk. Therefore, the reserve balances provide banks with an important buffer stock of liquid assets.

To overseas central banks. These are deposits of sterling (or euros in the case of the ECB) made by overseas authorities as part of their official reserves and/or for purposes of intervening in the foreign exchange market in order to influence the exchange rate of their currency. We will examine exchange rates in Chapter 13.

It operates the government's monetary policy

Monetary policy. The Bank of England's Monetary Policy Committee (MPC) sets the Bank Rate at its regular meetings. This nine-member committee consists of four experts appointed by the Chancellor of the Exchequer and four senior members of the Bank of England, plus the Governor in the chair.

By careful management of the liquidity of the financial system the Bank of England aims to keep market interest rates in line with Bank Rate. It is able to do this through operations in the money markets. These are known as **open-market operations** (**OMOs**).

If shortages of liquidity are driving up short-term interest rates above the desired level, the Bank of England purchases securities (gilts and/or Treasury bills) on the open market: e.g. through reverse repos (a repo to the banks). This releases liquidity into the financial system and puts downward pressure on interest rates. Conversely, if excess liquidity is driving

Definition

Open-market operations The sale (or purchase) of government securities in the open market which aim to reduce (or increase) the money supply and thereby affect interest rates.

down interest rates, the Bank of England will sell more securities. When these are purchased, this will reduce banks' reserves and thereby put upward pressure on interest rates.

As the 2007–8 global financial crisis unfolded, it became increasingly difficult for the Bank to meet its monetary policy objectives while maintaining financial stability. New policies were thus adopted and short-term open-market operations ceased. The key priority was now ensuring sufficient liquidity and so the focus switched to longer-term OMOs.

March 2009 saw the Bank begin a programme of **quantitative easing** (**QE**) (see Box 13.5). The aim was to increase the amount of money in the financial system and thereby stimulate bank lending and hence aggregate demand. QE involved the Bank creating electronic money and using it to purchase assets, mainly government bonds, predominantly from non-deposit-taking financial institutions, such as unit trusts, insurance companies and pension funds. These institutions would then deposit the money in banks, which could lend it to businesses and consumers for purposes of spending.

It provides liquidity, as necessary, to banks

Financial institutions engage in maturity transformation. While most customer deposits can be withdrawn instantly, financial institutions will have a variety of lending commitments, some of which span many years. Hence, the Bank of England acts as a 'liquidity backstop' for the banking system. It attempts to ensure that there is always an adequate supply of liquidity to meet the legitimate demands of depositors in banks.

Banks' reserve balances provide them with some liquidity insurance. However, the Bank of England needs other means by which to provide both individual banks and the banking system with sufficient liquidity. The financial crisis, for instance, saw incredible pressure on the aggregate liquidity of the financial system. The result is that the UK has three principal insurance facilities:

Index long-term repos (ILTRS). Each month the Bank of England provides MFIs with reserves for a six-month period secured against collateral and indexed against the Bank Rate. Financial institutions can borrow reserves against different levels of collateral. These levels reflect the quality and liquidity of the collateral. The reserves are distributed through an auction where financial institutions indicate, for their particular level of collateral, the number of basis points over the Bank Rate (the 'spread') they are prepared to pay. The resulting equilibrium interest rate, paid by all those borrowing, is that which balances the demand from MFIs with the supply of reserves made available. The frequency of the auctions can be varied to ensure sufficient liquidity.

Discount window facility (DWF). This on-demand facility allows financial institutions to borrow government bonds (gilts) for 30 days against different classes of (less liquid) collateral. They pay a fee to do so. The size of the fee is determined by both the type and quantity of collateral being traded. The gilts can then be used in repo operations as a means of securing liquidity. Financial institutions can look to roll over the gilts obtained from the DWF beyond the normal 30 days if they are still short of liquidity.

Contingent term repo facility (CTRP). This is a facility which the Bank of England can activate in exceptional circumstances, such as it did in March 2020 in response to the disruption to financial markets resulting from the COVID-19 pandemic. As with the ILTRS, financial institutions can obtain liquidity secured against different levels of collateral through an auction. However, the terms, including the maturity of the funds, are intended to be more flexible.

It oversees the activities of banks and other financial institutions

The Bank of England requires all recognised banks to maintain adequate liquidity: this is called **prudential control**.

In May 1997, the Bank of England ceased to be responsible for the detailed supervision of banks' activities. This responsibility passed to the Financial Services Authority (FSA). But, the financial crisis of the late 2000s raised concerns about whether the FSA, the Bank of England and HM Treasury were sufficiently watchful of banks' liquidity and the risks of liquidity shortage. Some commentators argued that a much tighter form of prudential control needed to be imposed.

In 2013, the FSA was wound up and a new regulatory framework came into force, with an enhanced role for the Bank of England.

- First, the Bank's *Financial Policy Committee (FPC)* was made responsible for *macro-prudential regulation*. This type of regulation takes a broader view of the financial system. It considers the resilience of the financial system to possible shocks and its propensity to create macroeconomic instability through excessive

Definitions

Quantitative easing (**QE**) When the central bank increases the monetary base through an open-market purchase of government bonds or other securities. It uses electronic money (reserve liabilities) created specifically for this purpose.

Prudential control The insistence by the Bank of England that banks maintain adequate liquidity.

credit creation. In ensuring compliance with the Basel III regulatory framework (see pages 313–7), it sets the countercyclical buffer and, if necessary, can adjust the operation of the domestic systemic risk buffer.

- Second, the prudential regulation of individual firms was transferred from the FSA to the *Prudential Regulation Authority (PRA),* a subsidiary of the Bank of England.
- Third, the *Financial Conduct Authority (FCA)* took responsibility for consumer protection and the regulation of markets for financial services. The FCA is an independent body accountable to HM Treasury.

It operates the government's exchange rate policy

The Bank of England manages the country's gold and foreign currency reserves on behalf of the Treasury. This is done through the **exchange equalisation account**. As we shall see in Chapter 15, by buying and selling foreign currencies on the foreign exchange market, the Bank of England can affect the exchange rate.

Pause for thought

1. *Would it be possible for an economy to function without a central bank?*
2. *What effect would a substantial increase in the sale of government bonds and Treasury bills have on interest rates?*

The role of the money markets

Money markets enable participants, such as banks, to lend to and borrow from each other. The financial instruments traded are short-term ones. As we have seen, central banks use money markets to exercise control over interest rates. However, these markets are also very important in widening the lending and borrowing opportunities for financial institutions.

We take the case of the London money market, which is normally divided into the 'discount' and 'repo' markets and the 'parallel' or 'complementary' market.

The discount and repo markets

The discount market. The **discount market** is the market for commercial or government bills. In the UK, government bills are known as Treasury bills and operations are conducted by the Debt Management Office, usually on a weekly basis. Treasury bills involve short-term lending, say for 1 or 3 months, which, in conjunction with their low default risk, make them highly liquid assets.

The term 'discount market' is used because the bills being traded are issued at a price below their face value: i.e. at a discount. On maturity the holder is paid the full face value and thus earns the difference between that and the discounted purchase price. The *rate* of discount on bills can be calculated by the size of the discount relative to the redemption value and is usually expressed as an annual percentage rate.

The discount market is also known as the 'traditional market' because it was the market in which many central banks traditionally used to supply central bank money to financial institutions. For instance, if the Bank of England wanted to increase liquidity in the banking system it could purchase from the banks Treasury bills which had yet to reach maturity. This process is known as **rediscounting**. The Bank of England would pay a price below the face value, thus effectively charging interest to the banks. The price could be set so that the 'rediscount rate' reflected the Bank Rate.

The repo market. The emergence of the repo market is a more recent development dating back in the UK to the 1990s. Repos have become an important source of wholesale funding for financial institutions. But they are also an important means by which central banks can affect the liquidity of the financial system both to implement monetary policy and to ensure financial stability.

By entering into a repo agreement, the Bank of England can buy securities, such as gilts, from the banks (thereby supplying them with money) on the condition that the banks buy them back at a fixed price and on a fixed date. The repurchase price will be above the sale price. The difference is the equivalent of the interest that the banks are being charged for having what amounts to a loan from the Bank of England. The repurchase price (and hence the 'repo rate') is set by the Bank of England to reflect the Bank Rate chosen by the MPC.

The Bank of England first began using repo operations to manage the liquidity of the financial system in 1997 when it undertook daily operations, with the repurchases of securities usually occurring two weeks after the initial sale. This system was refined so that in 2006 operations became weekly and the repurchase period typically shortened to one week.

However, the financial crisis caused the Bank to modify its repo operations to manage liquidity both for purposes of monetary policy and to ensure financial

Definitions

Exchange equalisation account The gold and foreign exchange reserves account in the Bank of England.

Discount market The market for corporate bills and Treasury bills whose initial price is below the redemption value.

Rediscounting bills of exchange Buying bills before they reach maturity.

stability. These changes included a widening of the securities eligible as collateral for loans and a suspension of short-term repo operations.

So central banks, like the Bank of England, are prepared to provide central bank money through the creation of reserves. Central banks are thus the ultimate guarantor of sufficient liquidity in the monetary system and, for this reason, are known as the **lender of last resort**.

The parallel money markets

Like repo markets, complementary or parallel money markets have grown rapidly in recent years. In part, this reflects the opening up of markets to international dealing, the deregulation of banking and money market dealing, and the desire of banks to keep funds in a form that can be readily switched from one form of deposit to another, or from one currency to another.

Examples of parallel markets include the markets for *certificates of deposit* (CDs), *foreign currencies markets* (dealings in foreign currencies deposited short term in the country) and *the interbank market*. We focus here on the important interbank market (details on other parallel markets can be found in Case Study 11.9 on the student website).

The interbank market. This involves wholesale loans from one bank to another from one day to up to several months. Banks with surplus liquidity lend to other banks, which then use this as the basis for loans to individuals and companies.

The rate at which banks lend to each other has a major influence on the other rates that banks charge. Interbank loans typically can be anything from overnight to 12 months and hence there is a series of rates corresponding to these different maturities.

In the UK, the interbank rate has traditionally been measured by the LIBOR (the London interbank offered rate), varying by the length of loan. (In the eurozone the interbank rate is known as the Euribor, with the weighted average of all overnight rates known as Eonia). Interbank rates are the cost of borrowing between banks and have to be reported by the banks. However, in 2012 there were allegations of false reporting by financial institutions of LIBOR rates. Because the LIBOR rate was a reference rate for financial products this has potentially serious implications for the interest rates being charged.

From April 2016 the Bank of England took over the administration of an alternative interbank benchmark rate. Known as SONIA, the Sterling Overnight Index Average measures the rate at which interest is paid on sterling unsecured loans of one business day. SONIA has become the new benchmark determining commercial interest rates.

Figure 11.3 shows how closely SONIA typically tracks the Bank Rate. The exception was during the financial crisis of the late 2000s, when financial markets were in turmoil and the Bank of England loosened monetary policy aggressively.

Definition

Lender of last resort The role of the Bank of England as the guarantor of sufficient liquidity in the monetary system.

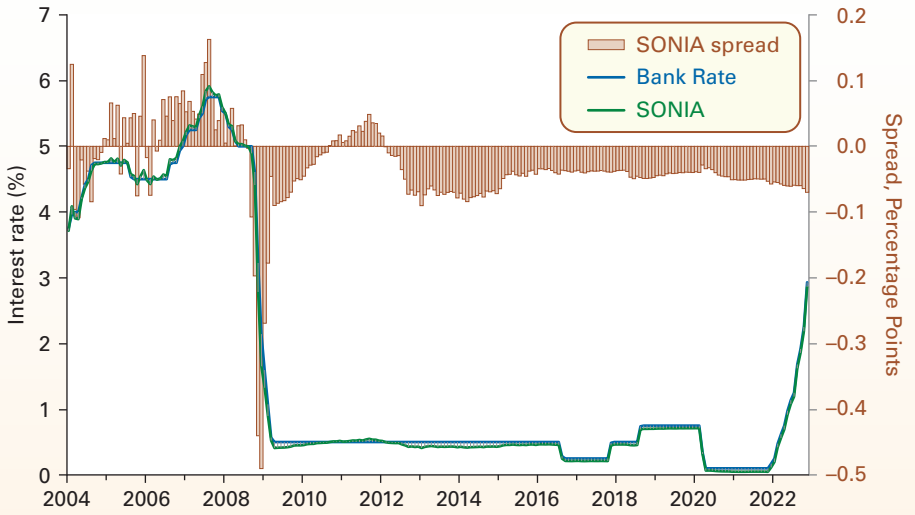
Pause for thought

Why should Bank of England determination of the rate of interest in the discount and repo markets also influence rates of interest in the parallel markets?

Recap

1. Central to the financial system are the retail and wholesale arms of banks. Between them they provide the following important functions: giving expert advice, channelling capital to areas of highest return, maturity transformation, risk transformation and the transmission of payments. During the financial crisis the systemic importance of some of these banks meant they had to be rescued by governments. They were seen as too important or too big to fail (TBF).
2. Banks' liabilities include both sight and time deposits. They also include certificates of deposit and repos. Their assets include: notes and coin, balances with the central bank, market loans, bills of exchange (Treasury bills and commercial bills), reverse repos, advances to customers (the biggest item – including overdrafts, personal loans, credit-card debt and mortgages) and investments (government bonds and interbank investments). In the years up to 2008 they had increasingly included securitised assets.
3. Banks aim to make profits, but they must also have a sufficient capital base and maintain sufficient liquidity. Liquid assets, however, tend to be relatively unprofitable and profitable assets tend to be relatively illiquid. Banks therefore need to keep a balance of profitability and liquidity in their range of assets, while also having adequate levels of capital to meet the demands of depositors and to cover potential losses from investments or if borrowers default on payments.
4. The regulatory framework provided by the Basel Committee on Banking Supervision has been strengthened following the financial crisis. The key features of Basel III include: increasing the level and quality of capital (capital adequacy), constraining leverage and hence the build-up of debt to fund activities, improving bank liquidity and limiting pro-cyclical lending behaviour. Additional capital and leverage buffers are applied to global systemically important banks (GSIBs).

Figure 11.3 SONIA and Bank Rate (monthly averages)



Note: The Sterling Overnight Index Average (SONIA) reflects the average interest rate that banks pay to borrow sterling overnight from other financial institutions

Source: Based on data from *Statistical Interactive Database,* series IUMABEDR and IUMASOIA (Bank of England) (data published November 2022)

5. The Bank of England is the UK's central bank. It issues notes; it acts as banker to the government, to banks and to various overseas central banks; it ensures sufficient liquidity for the financial sector; it operates the country's monetary and exchange rate policy.
6. The money market is the market in short-term deposits and loans. It consists of the discount and repo markets and the parallel money markets.
7. Through repos the Bank of England provides liquidity to the banks at the rate of interest chosen by the Monetary Policy Committee (Bank Rate). It is always prepared to lend in this way in order to ensure adequate liquidity in the economy. The financial crisis saw the Bank of England adapt its operations in the money market and introduce new mechanisms for providing liquidity insurance, including the Discount Window Facility and longer-term repos.
8. The parallel money markets consist of various markets in short-term finance between various financial institutions.

11.3 THE SUPPLY OF MONEY

How is it measured and what determines its size?

If money supply is to be monitored and possibly controlled, it is obviously necessary to measure it. But what should be included in the measure? Here we need to distinguish between the *monetary base* and *broad money*.

The **monetary base** (or 'high-powered money' or 'narrow money') consists of cash (notes and coin) in circulation outside the central bank.[1] In 1970, the stock of notes and coins in circulation in the UK was around £4 billion, equivalent to 7 per cent of annual GDP. By 2022 this had grown to around £100 billion, but equivalent to only about 4 per cent of annual GDP.

KI 32 p 239

Definition

Monetary base Notes and coin outside the central bank.

[1]Before 2006, there used to be a measure of narrow money called 'M0'. This included cash in circulation outside the Bank of England and banks' non-interest bearing 'operational balances' in the Bank of England, with these balances accounting for a tiny proportion of the whole. Since 2006, the Bank of England has allowed banks to hold interest-bearing reserve accounts, which are much larger than the former operational balances. The Bank of England thus decided to discontinue M0 as a measure and focus on cash in circulation as its measure of the monetary base.

But the monetary base gives us a very poor indication of the effective money supply, since it excludes the most important source of liquidity for spending: namely, bank deposits – and most spending from these deposits is not by cash, but by debit card, direct debit, cheques, etc. The problem, however, is which deposits to include. We need to answer three questions:

- Should we include just sight deposits, or time deposits as well?
- Should we include just retail deposits, or wholesale deposits as well?
- Should we include just bank deposits, or building society (savings institution) deposits as well?

In the past there has been a whole range of measures, each including different combinations of these accounts. However, financial deregulation, the abolition of foreign exchange controls and the development of computer technology have led to huge changes in the financial sector throughout the world. This has led to a blurring of the distinctions between different types of account. It has also made it very easy to switch deposits from one type of account to another. For these reasons, the most usual measure that countries use for money supply is **broad money**, which in most cases includes both time and sight deposits, retail and wholesale deposits, and bank and building society (savings institution) deposits.

In the UK this measure of broad money is known as M4. In most other European countries and the USA, it is known as M3. There are, however, minor differences between countries in what is included.

In 1970, the stock of M4 in the UK was around £25 billion, equivalent to 48 per cent of annual GDP. By 2022 this had grown to £3.0 trillion, equivalent to about 130 per cent of annual GDP.

KI 32 p 239

As we have seen, bank deposits of one form or another constitute by far the largest component of (broad) money supply. To understand how money supply expands and contracts, and how it can be controlled, it is thus necessary to understand what determines the size of bank deposits. Banks can themselves expand the amount of bank deposits, and hence the money supply, by a process known as 'credit creation'.

The creation of credit

To illustrate this process in its simplest form, assume that banks have just one type of liability – deposits – and two types of asset – balances with the central bank (to achieve liquidity) and advances to customers (to earn profit).

Definition

Broad money Cash in circulation plus retail and wholesale bank and building society deposits.

Table 11.2 Banks' original balance sheet

Liabilities	£bn	Assets	£bn
Deposits	100	Balances with the central bank	10
		Advances	90
Total	100	Total	100

Banks want to achieve profitability while maintaining sufficient liquidity. Assume that they believe that sufficient liquidity will be achieved if 10 per cent of their assets are held as balances with the central bank. The remaining 90 per cent will then be in advances to customers. In other words, the banks operate a 10 per cent liquidity ratio.

Assume initially that the combined balance sheet of the banks is as shown in Table 11.2. Total deposits are £100 billion, of which £10 billion (10 per cent) are kept in balances with the central bank. The remaining £90 billion (90 per cent) are lent to customers.

Now assume that the government spends more money – £10 billion, say, on roads or education. It pays for this with cheques drawn on its account with the central bank. The people receiving the cheques deposit them in their banks. Banks return these cheques to the central bank and their balances correspondingly increase by £10 billion. The combined banks' balance sheet now is shown in Table 11.3.

But this is not the end of the story. Banks now have surplus liquidity. With their balances in the central bank having increased to £20 billion, they now have a liquidity ratio of 20/110, or 18.2 per cent. If they are to return to a 10 per cent liquidity ratio, they need only retain £11 billion as balances at the central bank (£11 billion/£110 billion = 10 per cent). The remaining £9 billion they can lend to customers.

Assume now that customers spend this £9 billion in shops using their debit cards. The balances in the

Table 11.3 The initial effect of an additional deposit of £10 billion

Liabilities	£bn	Assets	£bn
Deposits (old)	100	Balances with the central bank (old)	10
Deposits (new)	10	Balances with the central bank (new)	10
		Advances	90
Total	110	Total	110

Table 11.4 The full effect of an additional deposit of £10 billion

Liabilities	£bn	Assets	£bn
Deposits (old)	100	Balances with the central bank (old)	10
Deposits (new: initial)	10	Balances with the central bank (new)	10
(new: subsequent)	90	Advances (old)	90
		Advances (new)	90
Total	200	Total	200

central bank of the customers' banks will duly be debited by £9 billion, but the balances in the central bank of the shopkeepers' banks will be credited by £9 billion: leaving *overall balances in the central bank unaltered*. There is still a surplus of £9 billion over what is required to maintain the 10 per cent liquidity ratio. The new deposits of £9 billion in the shopkeepers' banks, backed by balances in the central bank, can thus be used as the basis for *further* loans. Ten per cent (i.e. £0.9 billion) must be kept back in the central bank, but the remaining 90 per cent (i.e. £8.1 billion) can be lent out again.

When the money is spent and again transferred between banks, this £8.1 billion will still remain as surplus balances in the central bank and can therefore be used as the basis for yet more loans. Again, 10 per cent must be retained and the remaining 90 per cent can be lent out. This process goes on and on until eventually the position is as shown in Table 11.4.

The initial increase in balances with the central bank of £10 billion has allowed banks to create new advances (and hence deposits) of £90 billion, making a total increase in money supply of £100 billion.

Pause for thought

If banks choose to operate with a 5 per cent liquidity ratio and receive an extra £100 million of cash deposits: (a) What is the size of the bank deposits multiplier? (b) How much will total deposits have expanded after the multiplier has worked through? (c) How much will total credit have expanded?

TC 14 p 249

This effect is known as the **bank deposits multiplier**. In this simple example, with a liquidity ratio of $^1/_{10}$ (i.e. 10 per cent), the bank deposits multiplier is 10. An initial increase in deposits of £10 billion allowed total deposits to rise by £100 billion. In this simple world, therefore, the bank deposits multiplier is the inverse of the liquidity ratio (L).

Bank deposits multiplier $= 1/L$

The creation of credit: the real world

In practice, the creation of credit is not as simple as this. There are three major complications.

Banks' liquidity ratio may vary

Banks may choose a different liquidity ratio. At certain times, banks may decide that it is prudent to hold a bigger proportion of liquid assets. For example, if banks are worried about increased risks of default on loans, they may choose to hold a higher liquidity ratio to ensure that they have enough to meet customers' needs. This was the case in the late 2000s when many banks became less willing to lend to other banks for fear of the other banks' assets containing sub-prime debt. Banks, as a result, hoarded cash and became more cautious about granting loans.

KI 15 p 80

On the other hand, there may be an upsurge in consumer demand for credit. Banks may be very keen to grant additional loans and thus make more profits, even though they have acquired no additional assets. They may simply go ahead and expand credit, and accept a lower liquidity ratio.

Customers may not want to take up the credit on offer. Banks may wish to make additional loans, but customers may not want to borrow. There may be insufficient demand. But will the banks not then lower their interest rates, thus encouraging people to borrow? Possibly; but if they lower the rate they charge to borrowers, they must also lower the rate they pay to depositors. But then depositors may switch to other institutions such as building societies.

Pause for thought

How will an increased mobility of savings and other capital between institutions affect this argument?

Banks may not operate a simple liquidity ratio

The fact that banks hold a number of fairly liquid assets, such as money at call, bills of exchange and certificates of deposit, makes it difficult to identify a simple liquidity ratio. If the banks use extra cash to buy such liquid assets, can they then use *these* assets as the basis for creating credit? It is largely up to banks'

Definition

Banks deposits multiplier The number of times greater the expansion of bank deposits is than the additional liquidity in banks that causes it: $1/L$ (the inverse of the liquidity ratio).

judgements on their overall liquidity position. In practice, therefore, the size of the bank deposits multiplier will vary and is thus difficult to predict in advance.

Some of the extra cash may be withdrawn by the public

If extra cash comes into the banking system, and as a result extra deposits are created, part of them may be held by households and non-bank firms (known in this context as the **non-bank private sector**) as cash outside the banks. In other words, some of the extra cash leaks out of the banking system. This will result in an overall multiplier effect that is smaller than the full bank deposits multiplier. This overall multiplier is known as the **money multiplier**. It is defined as the *change* in total money supply expressed as a proportion of the *change* in the monetary base that caused it: $\Delta M_s/\Delta M_b$ (where M_s is total broad money supply and M_b is the monetary base).

TC 14 p 249

The broad money multiplier in the UK

TC 14 p 249

In the UK, the principal money multiplier measure is the broad money multiplier. This is given by $\Delta M4/\Delta M_b$, where M_b in this case is defined as cash in circulation with the public and in banks' interest-bearing deposits (reserve accounts) at the Bank of England.

Another indicator of the broad money multiplier is simply the ratio of the *level of* (as opposed to change in) M4 relative to the *level of* cash in circulation with the public and banks' reserve accounts at the central bank. This 'levels' relationship is shown in Figure 11.4 and helps us to analyse the longer-term relationship between the stocks of broad money and the monetary base.

From Figure 11.4 we can see how broad money grew rapidly relative to the monetary base during the 1980s and into the early 1990s. From the early 1990s to the mid-2000s, the level of M4 relative to the monetary base fluctuated in a narrow range.

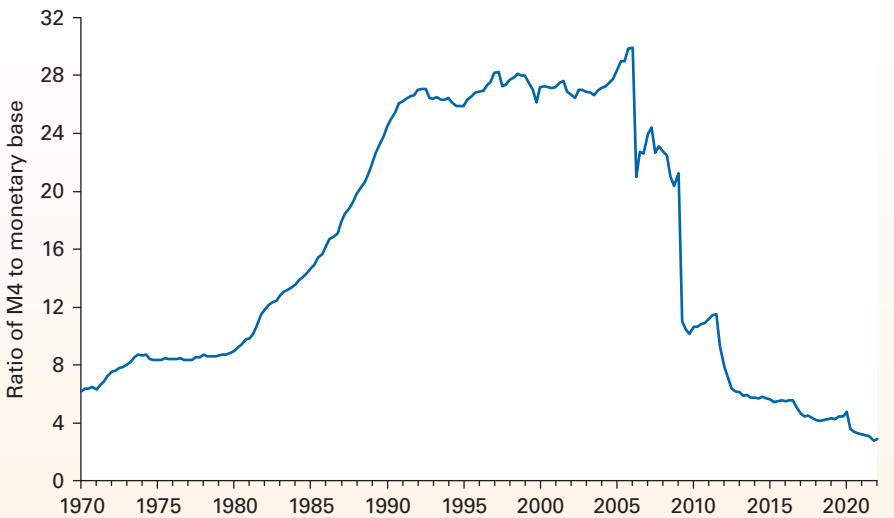
From May 2006 the Bank of England began remunerating banks' reserve accounts at the official Bank Rate. This encouraged banks to increase their reserve accounts at the Bank of England and led to a sharp fall in the broad money multiplier. It then declined further in 2009 and in 2011–12, coinciding with the Bank of England's programme of asset purchases (quantitative easing) in response to the global financial crisis. The increase in the monetary base, however, did not lead to the same percentage increase in broad money, as banks were cautious about lending and chose to keep higher reserves. Further asset purchases were made in 2016, following the EU referendum vote, and in 2020–21 in response to the pandemic. Again the broad money multiplier fell, albeit less dramatically. The policy of quantitative easing is discussed further in Chapter 13.

Definitions

Non-bank private sector Households and non-bank firms. In other words, everyone in the country other than banks and the government (central and local).

Money multiplier The number of times greater the expansion of money supply (M_s) is than the expansion of the monetary base (M_b) that caused it: $\Delta M_s/\Delta M_b$.

Figure 11.4 UK broad money multiplier



Note: Quarterly figures for the monetary base (denominator) are taken from monthly data series using the last month of the quarter

Source: Based on series LPQAUYN (M4), LPMAVAE (M0) until 2006, LPMBL22 (reserves) and LPMAVAB (notes and coin) from 2006, *Statistical Interactive Database*, Bank of England (data published 1 July 2022, seasonally adjusted except for reserves)

In the next section we look at factors which help explain movements in the money multiplier and changes in the money supply.

Pause for thought

Which would you expect to fluctuate more, the money multiplier ($\Delta M_s/\Delta M_b$), or the simple ratio, M_s/M_b?

What causes money supply to rise?

Money supply can rise for a number of reasons. We consider five sets of circumstances which can cause the money supply to *rise*.

Central bank action

The central bank may decide that the stock of money is too low and that this is keeping up interest rates and holding back spending in the economy. In such circumstances, it may choose to create additional money.

As we saw above, this was the case following the global financial crisis and the COVID-19 pandemic, when central banks around the world embarked on programmes of *quantitative easing*. QE involves the central bank creating electronic (narrow) money and using it to purchase assets, mainly government bonds. When the recipients of the money (mainly non-bank financial institutions) deposit it in banks, the banks can then lend it to businesses and consumers for purposes of spending and, through the bank deposits multiplier, broad money supply would increase.

TC 14 p 249

As we can see from Figure 11.5, however, this was not enough to prevent UK broad money supply contracting across much of the first half of the 2010s.

An inflow of funds from abroad

When sterling is used to pay for UK exports and is deposited in UK banks by the exporters, credit can be created on the basis of it. This leads to a multiplied increase in the domestic money supply.

TC 14 p 249

The money supply will also expand if depositors of sterling in banks overseas then switch these deposits to banks in the UK. This is a direct increase in the money supply. In an open economy like the UK, movements of sterling and other currencies into and out of the country can be very large. This can lead to large fluctuations in the money supply.

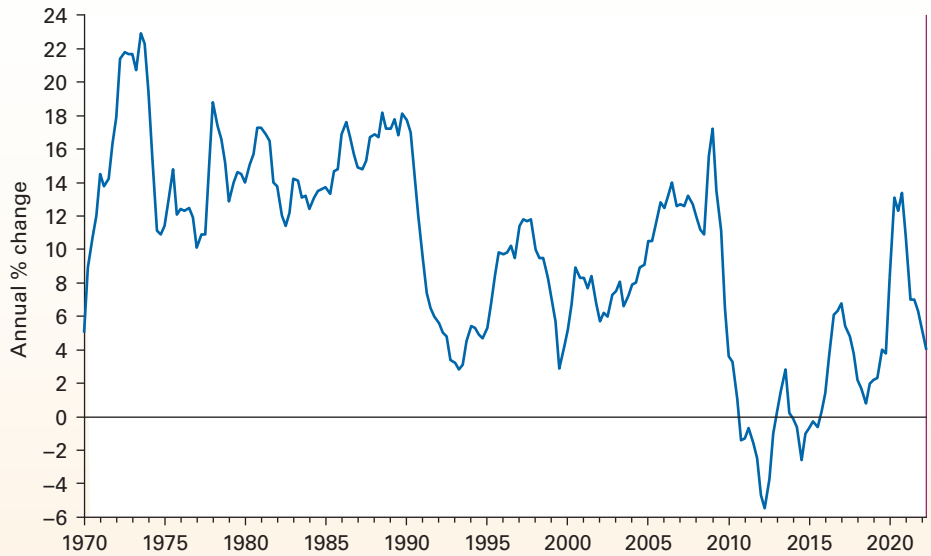
A public-sector deficit

A public-sector deficit is the difference between public-sector expenditure and public-sector receipts. To finance this deficit, the government has to borrow money by selling interest-bearing securities (Treasury bills and gilts). The precise amount of money the public sector requires to borrow in any one year is known in the UK as the **public-sector net cash requirement** (**PSNCR**).

Definition

Public-sector net cash requirement (PSNCR) A measure of the (annual) deficit of the public sector, and thus the amount it must borrow.

Figure 11.5 Annual rate of growth of M4



Source: *Statistical Interactive Database* (Bank of England), Series LPMVQJW (data published 30 September 2022, seasonally adjusted)

We will focus on public-sector spending and taxation in Chapter 13.

In general, the bigger the public sector's deficit, the greater the growth in the money supply will be. Just how the money supply will be affected, however, depends on who buys the securities

Consider first the case where government securities are purchased by the non-bank private sector. The money supply will remain unchanged. When people or firms buy the bonds or bills, they will draw money from their banks. When the government spends the money, it will be redeposited in banks. There is no increase in money supply. It is just a case of existing money changing hands.

This is not the case when the securities are purchased by the banking sector, including the central bank. Consider the purchase of Treasury bills by commercial banks: there will be a multiplied expansion of the money supply. The reason is that, although banks' balances at the central bank will go down when the banks purchase the bills, they will go up again when the government spends the money. In addition, the banks will now have additional liquid assets (bills), which can be used as the basis for credit creation.

The government could attempt to minimise the boost to money supply by financing the deficit through the issue of debt instruments of longer maturity (gilts), since these are less liquid.

Pause for thought

Identify the various factors that could cause a fall in the money supply.

The above reasons for an expansion of broad money supply (M4) are reasons why the monetary base itself might expand. The following two reasons focus on how the money supply can increase if more *credit* is created for a given monetary base. These reasons therefore would see a rise in the money multiplier, such as the UK saw in the 1980s and then again in the first half of the 2000s (see Figure 11.4).

Banks choose to hold a lower liquidity ratio

If banks collectively choose to hold a lower liquidity ratio, they will have surplus liquidity. The banks have tended to choose a lower liquidity ratio over time because of the increasing use of direct debits and debit-card and credit-card transactions.

Surplus liquidity can be used to expand advances, which will lead to a multiplied rise in broad money supply (e.g. M4).

An important trend up to the late 2000s was the growth in *interbank lending*. Short-term loans to other banks (including overseas banks) may be used by a bank as the basis for expanding loans and thereby starting a chain of credit creation. But although these assets are liquid to an *individual bank*, they do not add to the liquidity of the banking system *as a whole*. By using them for credit creation, the banking system is operating with a lower *overall* liquidity ratio.

This was a major element in the banking crisis of 2008. By operating with a collectively low liquidity ratio, banks were vulnerable to people defaulting on debt, such as mortgages. The problem was compounded by the holding of sub-prime debt in the form of securitised assets. Realising the vulnerability of other banks, banks became increasingly unwilling to lend to each other. The resulting decline in interbank lending reduced the amount of credit created and so depressed the money supply (see Figure 11.5).

Pause for thought

What effects do debit cards and cash machines (ATMs) have on (a) banks' prudent liquidity ratios; (b) the size of the bank deposits multiplier?

The non-bank private sector chooses to hold less cash

Households and firms may choose to hold less cash. Again, the reason may be a greater use of cards, direct debits, etc. This means that a greater proportion of the cash base will be held as deposits in banks rather in people's wallets, purses or safes outside banks. The extra cash deposits allow banks to create more credit.

The flow of funds equation

KI 23 p 173

A **flow-of-funds equation** can be used to analyse the contributions to money growth made by the public-sector's budget balance, i.e. the balance between pubic-sector spending and taxation, lending by financial institutions and flows of funds to and from abroad.

Definition

Flow-of-funds equation The various items making up an increase (or decrease) in money supply.

The following flow-of-funds equation is the one most commonly used in the UK, that for M4. It consists of four components:

ΔM4	*equals*	PSNCR	(Item 1)
	minus	Sales of public-sector debt to (or plus purchases of public-sector debt from) the non-bank private sector	(Item 2)
	plus	Banks' and building societies' sterling net lending to the UK private sector	(Item 3)
	plus	External effect	(Item 4)

Public-sector borrowing (item 1) will lead to a direct increase in the money supply, but not if it is funded by selling bonds and bills to the non-bank private sector. Such sales (item 2) have therefore to be subtracted from the public-sector net cash requirement (PSNCR). But conversely, if the government buys back old bonds from the non-bank private sector, this will further increase the money supply.

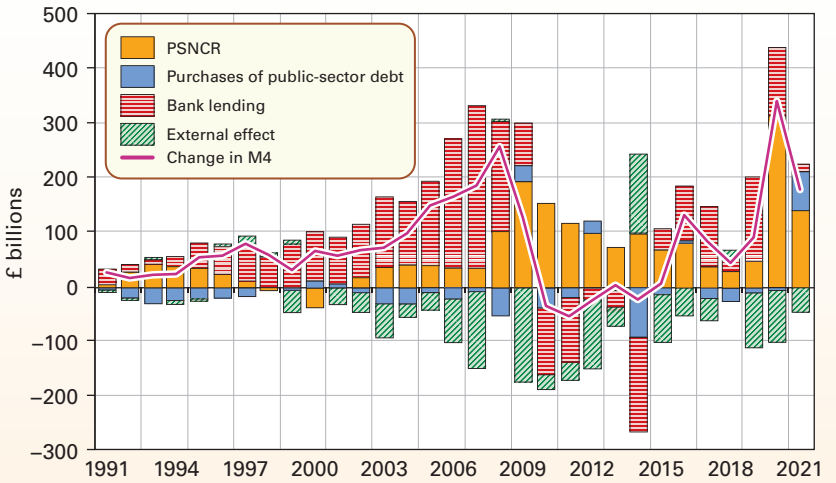
The initial increase in liquidity from the sale of government securities to the banking sector is given by item 1. This increase in their liquidity will enable banks to create credit. To the extent that this extra lending is to the UK private sector (item 3), money supply will increase, and by a multiple of the initial increase in liquidity (item 1).

Bank lending may also increase (item 3) even if there is no increase in liquidity or even a reduction in liquidity (item 1 is zero or negative). This happens if banks respond to increases in the demand for loans by accepting a lower liquidity ratio, or if, through securitisation and other forms of secondary marketing, individual banks gain extra liquidity from each other, even though there is no total increase in liquidity in the banking system. Item 3 will be reduced if banks either choose to, or because of regulatory requirements, hold more capital.

Finally, if there is a net inflow of funds from abroad (item 4), this too will increase the money supply.

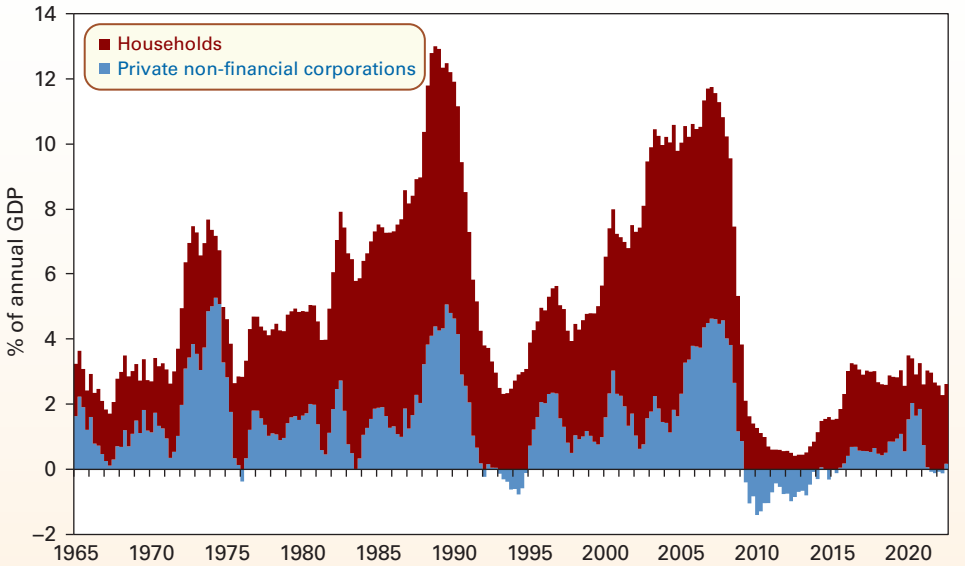
Figure 11.6 shows the components of changes in broad money (M4) in the UK since 1991. It illustrates the contribution from bank lending to the strong growth of the money supply during the 2000s, up to the financial crisis. However, the curbing of lending by banks following the financial crisis, and then the enhanced capital adequacy requirements of the new Basel framework (see pages 313–16), dampened monetary growth. This helped to offset the otherwise expansionary impact on the money supply of the large public-sector deficits in the first half of the 2010s. A resurgence in lending along with still large public-sector deficits helped to boost the growth in broad money in the remainder of the 2010s. Then, in the early 2020s, the fiscal measures in response to the COVID-19 pandemic and subsequent global economic turbulence led to historically large budget deficits, boosting monetary growth.

Figure 11.6 Components of changes in M4



Source: Based on data in Bankstats (Bank of England), Table A3.2 (data published 29 March 2022)

Figure 11.7 Annual flows of credit from MFIs



Sources: Statistical Interactive Database, Bank of England, series LPQVWNV and LPQVWNQ (data published 31 October 2022, seasonally adjusted) and series YBHA, Office for National Statistics

Credit cycles

As we have seen, an analysis of monetary growth shows the importance of bank lending. A closer analysis of the flows of credit extended by financial institutions shows that they can vary considerably from period to period. This variation is consistent with the idea of **credit cycles**. Credit cycles can therefore help to explain some of the patterns we see in the money supply. They may also be important in shaping the business cycle or how economies adjust to economic shocks.

TC 3 p 10

Credit cycles can be seen readily in Figure 11.7. The chart shows the annual flows of credit from monetary financial institutions (MFIs) to households and private non-financial corporations in the UK from 1970. The flows are measured relative to GDP, allowing us to make better comparisons of the magnitude of credit flows over time.

As you can see, there were especially large flows of credit from MFIs in the late-1980s and the early/mid-2000s. In each case, this was then followed by a period of significantly weaker credit growth and, for non-financial corporations, even net repayments of outstanding lending.

The financial crisis has helped to stimulate the interest among economists and policy makers in the causes of credit cycles, and their relationship with aggregate demand and the growth of national income. In Chapter 10, we saw how credit conditions might become dependent on the current macroeconomic environment if it affects banks' lending behaviour. If so, then lending can be pro-cyclical and so amplify the business cycle through a *financial accelerator effect*.

TC 14 p 249

Box 11.4 considers another perspective on credit cycles: *The Financial Instability Hypothesis*. The argument here is that credit cycles are the primary source of the business cycle.

The relationship between money supply and the rate of interest

Simple monetary theory often assumes that the supply of money is totally independent of interest rates. The money supply is **exogenous**. This is illustrated in Figure 11.8(a). The supply of money is assumed to be determined by the government or central bank ('the authorities'): what the authorities choose it to be, or what they allow it to be by their choice of the level and method of financing public-sector borrowing.

In practice, however, even if the monetary base were to be tightly controlled by the central bank

Definitions

Credit cycle The expansion and contraction of credit flows over time.

Exogenous money supply Money supply that does not depend on the demand for money but is set by the authorities (i.e. the central bank or the government).

BOX 11.4 MINSKY'S FINANCIAL INSTABILITY HYPOTHESIS

EXPLORING ECONOMICS

Are credit cycles inevitable?

Destabilising financial cycles

Hyman Minsky (1919–1996) was an American economist, born of Belarusian parents. He is known for his work on understanding the relationship between the financial system and the macroeconomy.

KI 4 p 10

His **financial instability hypothesis** proposes that the volume of credit flows goes through stages which are ultimately destabilising for the economy.

In the expansionary stage, flows of credit increase rapidly. This feeds the growth in aggregate demand and, as incomes expend, so people and businesses borrow more, further stimulating the growth in demand.

But, as people and businesses take on more and more debt, so their borrowing becomes unsustainable. This then results in a period of consolidation when they seek (or are required by banks) to reduce their debts. Spending growth is thus subdued and the growth in aggregate demand ends or slows down.

KI 32 p 239

Minsky argued that financial cycles are an inherent part of the economic cycle and are the primary source of fluctuations in real GDP. He argued that the accumulation of debt by people and businesses is not only pro-cyclical, but destabilising.

While Minsky himself focused on the accumulation of debt by businesses, the role of the mortgage market in generating unsustainable stocks of household debt during the 2000s (see Box 11.2) has meant that his ideas are now frequently applied across the whole of the private sector.

From tranquillity to bust

The extension of credit by financial institutions can be seen to go through three different stages: financial tranquillity, financial fragility and financial bust. During each of these stages both the credit criteria of banks and the ability of borrowers to afford their debts vary.

Minsky argued that credit flows will tend to increase during a period of sustained growth. This causes banks and investors to develop a heightened euphoria and confidence in the economy and in the returns of assets. Economists today refer to this as **irrational exuberance**, which results in people and businesses taking on bigger debts to acquire assets. These debts increasingly stretch their financial well-being. A point is reached, perhaps triggered by an economic shock or a tightening of economic policy, when the euphoria stops and then confidence is replaced with pessimism. This moment is now commonly referred to as a **Minsky moment**.

The result of a Minsky moment is that lenders reduce their lending, and people attempt to increase their net worth (i.e. reduce debts or increase savings) to ensure their financial well-being. However, this causes a decline in spending and in national income. The collapse in aggregate demand from financial distress therefore leads to a *balance-sheet recession*. This is exacerbated if the sale of assets, such as property or shares, causes their value to fall. The paradoxical reduction of net worth is known as the **paradox of debt**.

KI 32 p 239

Minsky believed that credit cycles are inevitable in a free-market economy. Hence, the authorities will need to act to moderate credit cycles. The significance given to macro-prudential regulation by policy makers (see Section 11.2) in the response to the financial crisis is recognition of the dangers posed to the economy by credit cycles. The hope is that the regulatory reforms will limit the amplification of the business cycle and provide greater financial resilience against shocks, such as the COVID-19 pandemic.

TC 7 p 26

A super cycle

While Minsky argued that the ingredients for economic volatility arising from financial instability are ever-present, some argue that other factors are needed for this instability to develop into a financial crisis. These factors may be part of a longer cycle of events. We could view the processes of financial deregulation and innovation that characterised the two to three decades leading up to the financial crisis as part of this longer cycle.

One interpretation of the financial crisis of the late 2000s is that it was the result of the interaction of the normal Minsky cycle (i.e. short-run variations in the accumulation of credit) with a longer cycle of events or a 'Minsky super-cycle'.

What demand-side and supply-side factors might be expected to influence the flows of net lending by financial institutions to the non-bank private sector?

Undertake an Internet news search for reference to the concept of a Minsky Moment. Summarise the findings from your search.

Definitions

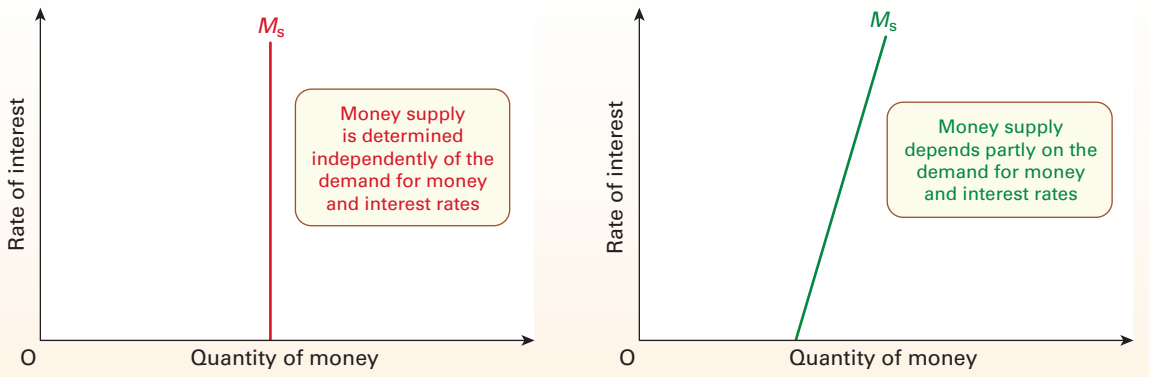
Financial instability hypothesis The theory that the economy goes through stages of credit accumulation which initially fuel aggregate demand but because of increasing financial distress eventually see a collapse in aggregate demand.

Irrational exuberance Where banks and other economic agents are over confident about the economy and/or financial markets and expect economic growth to remain stronger and/or asset prices to rise further than warranted by evidence. The term is associated with the economist Robert Shiller and his book *Irrational Exuberance* (2000) and with the former US Federal Reserve Chairman, Alan Greenspan.

Minsky moment A turning point in a credit cycle, where a period of easy credit and rising debt is replaced by one of tight credit and debt consolidation.

Paradox of debt (or paradox of deleveraging) The paradox that one individual can increase his or her net worth by selling assets, but if this is undertaken by a large number of people aggregate net worth declines because asset prices fall.

Figure 11.8 The supply of money curve



(which it is not), it would be very hard to have a precise control of broad money. More complex models, therefore, and especially Keynesian models, assume that money supply is **endogenous**: that it depends on the demand for money. The argument is that higher money demand will result in higher interest rates and in higher levels of money supplied. The result is a movement along an upward-sloping money supply curve, as in Figure 11.8(b). The reasons for this are as follows:

Definition

Endogenous money supply Money supply that is determined (at least in part) by the demand for money.

Recap

1. Money supply can be defined in a number of different ways, depending on what items are included. A useful distinction is between narrow money and broad money. Narrow money includes just cash and possibly banks' balances at the central bank. Broad money also includes deposits in banks and possibly various other short-term deposits in the money market. In the UK, M4 is the preferred measure of broad money. In the eurozone it is M3.
2. Bank deposits expand through a process of credit creation. If banks' liquid assets increase, they can be used as a base for increasing loans. When the loans are redeposited in banks, they form the base for yet more loans, and thus a process of multiple credit expansion takes place. The ratio of the increase of money to an expansion of the liquidity base is called the 'bank deposits multiplier'. It is the inverse of the liquidity ratio.
3. In practice, it is difficult to predict the precise amount by which money supply will expand if there is an increase in banks' liquidity. The reasons are that banks may choose to hold a different liquidity ratio; customers may not take up all the credit on offer; there may be no simple liquidity ratio, given the range of relatively liquid assets; and some of the extra cash may leak away into extra cash holdings by the public.
4. (Broad) money supply will rise if (a) banks choose to hold a lower liquidity ratio and thus create more credit for an existing amount of liquidity; (b) the non-bank private sector chooses to hold less cash; (c) the government runs a deficit and some of it is financed by borrowing from the banking sector; (d) there is an inflow of funds from abroad.
5. The flow-of-funds equation shows the components of any change in money supply. A rise in money supply equals the public-sector net cash requirement (PSNCR) *minus* sales of public-sector debt to the non-bank private sector, *plus* banks' lending to the private sector (less increases in banks' capital), *plus* inflows of money from abroad.
6. Bank lending exerts an important influence on monetary growth. Credit flows fluctuate in a way that is consistent with a credit cycle. The financial crisis has heightened interest in why a credit cycle arises and its impact on the macroeconomic environment.
7. Simple monetary theory assumes that the supply of money is independent of interest rates. In practice, a rise in interest rates (in response to a higher demand for money) will often lead to an increase in money supply, though this may be affected by the confidence of banks.

- *Banks accommodating an increase in the demand for credit*. Increases in money supply may occur as a result of banks expanding credit in response to the demand for credit. This may depend on banks having surplus liquidity, or being willing to operate with lower liquidity, or obtaining additional liquidity through secondary marketing of otherwise illiquid assets or from the central bank through repos. Higher demand for credit will drive up interest rates, making it more profitable for banks to supply more credit.
- *Depositors switch to less liquid deposits*. Higher interest rates may encourage depositors to switch their deposits from sight accounts (earning little or no interest) to time accounts. Since money is less likely to be withdrawn quickly from time accounts, banks may feel the need to hold less liquidity, and therefore may decide to increase credit, thus expanding the money supply.
- *Inflows of funds from abroad*. Higher interest rates attract deposits from overseas. A net inflow of funds from abroad, other things being equal, increases the money supply.

Some economists go further still. They argue that money supply is not only endogenous, but the 'curve' is effectively horizontal; money supply expands passively to match the demand for money.

It is likely, however, that the shape will vary with the confidence of banks. In periods of optimism, banks may be willing to expand credit to meet the demand from customers. In periods of pessimism, such as that following the financial crisis, banks may be unwilling to grant credit when customers seek it.

Pause for thought

What might the money supply curve look like in a period of irrational exuberance?

11.4 THE DEMAND FOR MONEY

How much money do we want to hold at any one time?

The demand for money refers to the desire to *hold* money: to keep your wealth in the form of money, rather than spending it on goods and services or using it to purchase financial assets such as bonds or shares. It is usual to distinguish three reasons why people want to hold their assets in the form of money:

The transactions motive. Since money is a medium of exchange, it is required for conducting transactions. But since people only receive money at intervals (e.g. weekly or monthly) and not continuously, they require to hold balances of money in cash or in current accounts.

The precautionary motive. Unforeseen circumstances can arise, such as a car breakdown. Thus individuals often hold some additional money as a precaution. Firms also keep precautionary balances because of uncertainties about the timing of their receipts and payments. If a large customer is late in making payment, a firm may be unable to pay its suppliers unless it has spare liquidity.

The assets or speculative motive. Money is not just a medium of exchange, it is also a means of storing wealth (see page 302). Keeping some or all of your wealth as money in a bank account has the advantage of carrying no risk. It earns a relatively small, but safe rate of return. Some assets, such as company shares or bonds, may earn you more on average, but there is a chance that their price will fall. In other words, they are risky.

KI 15 p 80

What determines the size of the demand for money?

What would cause the demand for money to rise? We now turn to examine the various determinants of the size of the demand for money (M_d). In particular, we will look at the role of the rate of interest. First, however, let us identify the other determinants of the demand for money.

Money national income. The more money people earn, the greater will be their expenditure and hence the greater the transactions demand for money. A rise in money ('nominal') incomes in a country can be caused either by a rise in real GDP (i.e. real output) or by a rise in prices, or by some combination of the two.

TC 5 p 17

The frequency with which people are paid. The less frequently people are paid, the greater the level of money balances that will be required to tide them over until the next payment.

Financial innovations. The increased use of credit cards, debit cards and cash machines, plus the advent of interest-paying current accounts, have resulted in

changes in the demand for money. The use of credit cards reduces both the transactions and precautionary demands. Paying once a month for goods requires less money on average than paying separately for each item purchased. Moreover, the possession of a credit card reduces or even eliminates the need to hold precautionary balances for many people.

On the other hand, the increased availability of cash machines, the convenience of debit cards and the ability to earn interest on current accounts have all encouraged people to hold more money in bank accounts. The net effect has been an increase in the demand for (broad) money.

TC 11 p 61

Speculation about future returns on assets. The assets motive for holding money depends on people's expectations. If they believe that share prices are about to fall on the stock market, they will sell shares and hold larger balances of money in the meantime. The assets demand, therefore, can be quite high when the price of securities is considered certain to fall. Some clever (or lucky) individuals anticipated the 2007–8 stock market decline. They sold shares and 'went liquid'.

Generally, the riskier such alternatives to money become, the more people will want to hold their assets as money balances in a bank or building society.

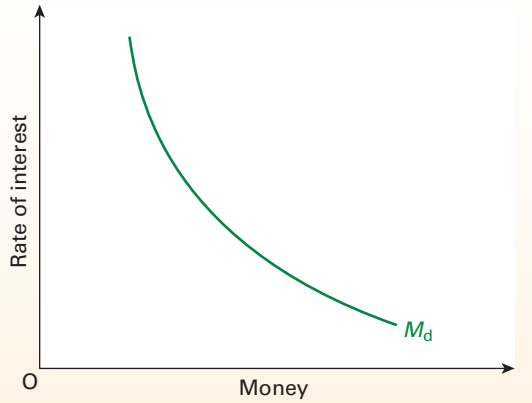
KI 15 p 80

People also speculate about changes in the exchange rate. If businesses believe that the exchange rate is about to appreciate (rise), they will hold greater balances of domestic currency in the meantime, hoping to buy foreign currencies with them when the rate has risen (since they will then get more foreign currency for their money).

The rate of interest (or rate of return) on assets. In terms of the operation of money markets, this is the most important determinant. It is related to the opportunity cost of holding money. The opportunity cost is the interest forgone by not holding higher interest-bearing assets, such as shares, bills or bonds. With most bank accounts today paying interest, this opportunity cost is less than in the past and thus the demand for money for assets purposes has increased.

But what is the relationship between money demand and the rate of interest? Generally, if rates of interest (or return) rise, they will rise more on shares, bills and bonds than on bank accounts. The demand for holding money in accounts will thus fall. The demand for money is thus inversely related to the rate of interest.

Figure 11.9 The demand for money



The demand for money curve

The demand for money curve with respect to interest rates is shown in Figure 11.9. It is downward sloping, showing that lower interest rates will encourage people to hold additional money balances.

A change in interest rates is shown by a movement along the demand for money curve. A change in any other determinant of the demand for money (such as national income or expectations about exchange rate movements) will cause the whole curve to shift: a rightward shift representing an increase in demand; a leftward shift representing a decrease.

Pause for thought

Which way is the demand for money curve likely to shift in each of the following cases? (a) Prices rise, but real incomes stay the same. (b) Interest rates abroad rise relative to domestic interest rates. (c) People anticipate that share prices are likely to fall in the near future.

Recap

1. The three motives for holding money are the transactions, precautionary and assets (or speculative) motives.
2. The demand for money will be higher, (a) the higher the level of money national income (i.e. the higher the level of real national income and the higher the price level), (b) the less frequently people are paid, (c) the greater the advantages of holding money in bank accounts, such as the existence of cash machines and the use of debit cards, (d) the more risky alternative assets become and the more likely they are to fall in value, and the more likely the exchange rate is to rise, and (e) the lower the opportunity cost of holding money in terms of interest forgone on alternative assets.
3. The demand for money curve with respect to interest rates is downward sloping.

11.5 EQUILIBRIUM

What effect do the demand and supply of money have on interest rates?

Equilibrium in the money market

TC 8 p 37

Equilibrium in the money market occurs when the demand for money (M_d) is equal to the supply of money (M_s). This equilibrium is achieved through changes in the rate of interest.

In Figure 11.10, assume that the demand and supply of money are given by M_s and M_d. The equilibrium rate of interest is i_e and the equilibrium quantity of money is M_e. But why?

If the rate of interest were above i_e, people would have money balances surplus to their needs. They would use these to buy shares, bonds and other assets. This would drive up the price of these assets. But the price of assets is inversely related to interest rates. The higher the price of an asset (such as a government bond), the less any given interest payment will be as a percentage of its price (e.g. £10 as a percentage of £100 is 10 per cent, but as a percentage of £200 is only 5 per cent). Thus a higher price of assets will correspond to lower interest rates.

As the rate of interest fell, so there would be a contraction of the money supply (a movement down along the M_s curve) and an increase in the demand for money balances, especially speculative balances (a movement down along the M_d curve). The interest rate would go on falling until it reached i_e. Equilibrium would then be achieved.

Similarly, if the rate of interest were below i_e, people would have insufficient money balances. They would sell securities, thus lowering their prices and raising the rate of interest until it reached i_e.

A shift in either the M_s or the M_d curve will lead to a new equilibrium quantity of money and rate of interest at the new intersection of the curves. For example, a rise in the supply of money will cause the rate of interest to fall, whereas a rise in the demand for money will cause the rate of interest to rise.

In practice, there is no one single interest rate. Rather equilibrium in the money markets will be where demand and supply of the various financial instruments separately balance. Generally, however, different interest rates tend to move roughly together as the overall demand for money and other liquid assets (or their supply) changes.

Figure 11.11 gives some examples of interest rates on various financial instruments. It shows how the various rates of interest generally move together, especially those with shorter maturities. As we saw in Section 11.2, the Bank of England conducts open-market operations to affect the general structure of the economy's interest rates and, in turn, the inflation rate. We can see how significant reductions to the policy rate (Bank Rate) following the financial crisis and the COVID-19 pandemic were mirrored by falls in other interest rates.

Figure 11.10 Equilibrium in the money market

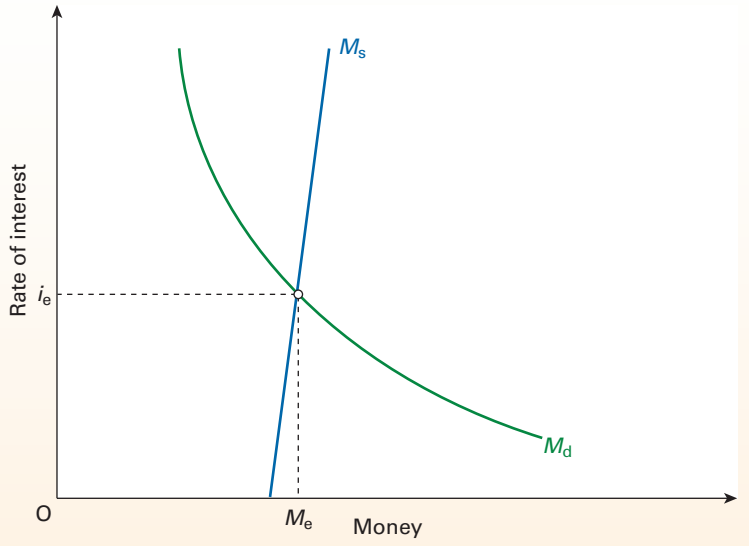
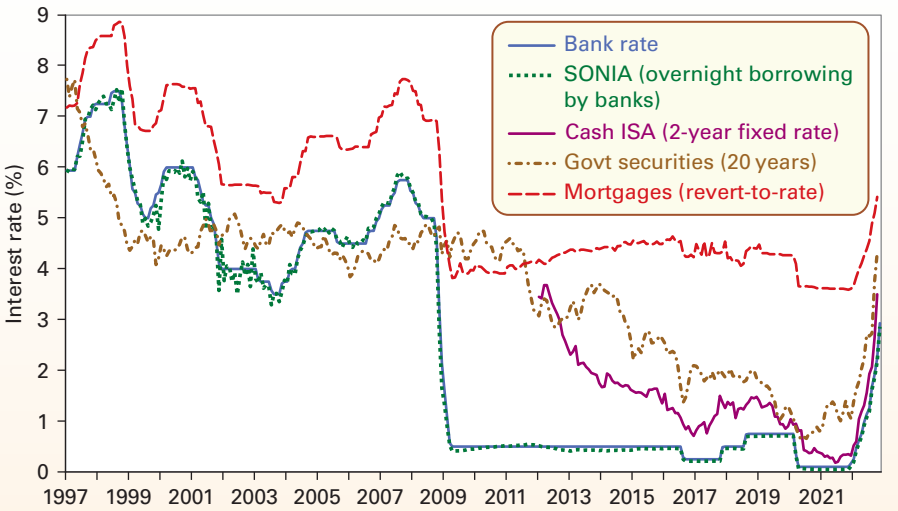


Figure 11.11 Selected interest rates (monthly averages)



Note: Government securities: zero coupon, nominal yields

Source: Based on data from *Statistical Interactive Database*, series IUMABEDR, IUMASOIA, IUMZID2, IUMALNZC, IUMTLMV, Bank of England (November 2022)

The link between the money and goods markets

TC 14 p 249

A rise in money supply will cause a rise in aggregate demand. This, in turn, cause a multiplied rise in national income. The size of the rise in real national income (i.e. output) will depend on the degree of spare capacity in the economy. In other words, as we saw in Section 10.2, it will depend on the slope of the short-run aggregate supply (*SRAS*) curve. The more elastic (flatter) the *SRAS* curve, the greater the rise in output relative to prices.

There are two principal ways in which a rise in money supply causes a rise in aggregate demand. The first is via changes in interest rates – known as the **interest-rate transmission mechanism**. The second is via changes in the exchange rate – known as the **exchange-rate transmission mechanism**.

The interest-rate transmission mechanism

The interest rate transmission mechanism is illustrated in the top part of Figure 11.12. It is a three-stage process.

- A rise in money supply leads to a surplus of money at the current rate of interest. This results in a fall in the equilibrium rate of interest. This would be illustrated by a rightward shift in the M_s line in Figure 11.10, with the rate of interest falling to the point where this new M_s curve crossed the M_d curve.
- This fall in the rate of interest then encourages firms to invest, since it is cheaper to borrow money to finance new buildings, machines, etc. A rise in investment would lead the injections (J) line in Figure 9.9 (see page 250) to move vertically upwards. It also encourages consumers to spend, since borrowing through credit cards and personal loans is now cheaper. At the same time, it discourages saving. This would lead to a downward shift in the withdrawals (W) line in Figure 9.9.
- The net effect of these changes in injections (increase) and withdrawals (fall) is a rise in aggregate demand.

The overall effect of a change in money supply on national income will depend on the size of the effect in each of the three stages. Economists keenly debate the magnitude of these effects . The debates essentially focus on the shapes of the curves in Figures 11.10 and 9.9 and whether they are likely to shift.

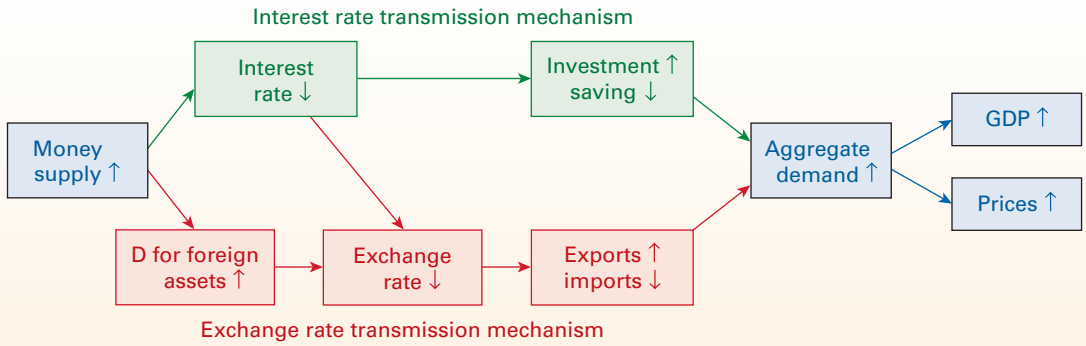
TC 10 p 49

Definitions

Interest-rate transmission mechanism How a change in money supply affects aggregate demand via a change in interest rates.

Exchange-rate transmission mechanism How a change in money supply affects aggregate demand via a change in exchange rates.

Figure 11.12 Monetary transmission mechanism



Pause for thought

If everybody believes that the rate of interest will rise, what effect will an increase in the money supply have on the rate of interest? Why was this question relevant in the late 2000s?

The exchange-rate transmission mechanism

Exchange rates are determined by the demand and supply of currencies. (We will examine this in detail in Chapter 15.) If the supply of sterling on the foreign exchange market (e.g. from importers in the UK wishing to buy foreign currencies in order to buy foreign goods) exceeds the demand (e.g. from foreign companies wishing to obtain sterling to buy UK exports), the exchange rate will fall ('depreciate'). For example, the pound might depreciate, say from €1.20 to €1.10. Conversely, if the demand for sterling exceeds the supply, the exchange rate will rise ('appreciate').

Changes in the money supply will not only affect interest rates, they will also affect the demand and supply of the currency and thus have an effect on exchange rates. The change in the exchange rate will then affect aggregate demand. This 'exchange-rate transmission mechanism' is illustrated in the bottom part of Figure 11.12.

Assume again that the money supply increases. This has the following effects:

- Part of the excess money balances is used to purchase foreign assets. This therefore leads to an increase in the supply of domestic currency coming on to the foreign exchange markets.
- As we have already seen, the excess supply of money in the domestic money market pushes down the rate of interest. This reduces the return on domestic assets below that on foreign assets. This, like the first effect, leads to an increased demand for foreign assets and thus an increased supply of domestic currency on the foreign exchange market. It also reduces the demand for domestic assets by those outside the country, and thus reduces the demand for the domestic currency. This causes the exchange rate to fall (depreciate).
- The fall in the exchange rate (e.g. from €1.20 to €1.10) means that people abroad have to pay less for a pound. This makes UK exports (an injection) cheaper and hence more are sold. People in the UK, by contrast, get less foreign currency for a pound. This makes imports (a withdrawal) more expensive and hence fewer are purchased.
- Again, the net effect of these changes in injections and withdrawals is a rise in aggregate demand.

In both cases, the rise in aggregate demand then leads to a multiplied rise in national income (GDP). This, however, as we shall see in Section 12.2, may be wholly or partly offset by a rise in prices.

TC 14 p 249

Pause for thought

What, do you think, determines the amount that real output rises as a result of a rise in the money supply? (We examine this in the next section.)

The relationship between the money and goods market is explored in more detail in Web Appendices 11.2 and 11.3, which look at the '*IS*/*LM*' and '*IS*/*MP*' models respectively.

Recap

1. Equilibrium in the money market is where the supply of money is equal to the demand. Equilibrium is achieved through changes in the interest rate.
2. The interest rate mechanism works as follows: a rise in money supply causes money supply to exceed money demand; interest rates fall; this causes investment (an injection) and consumer spending to rise and saving (a withdrawal) to fall; this causes a multiplied rise in national income.
3. The exchange rate mechanism works as follows: a rise in money supply causes interest rates to fall; the rise in money supply plus the fall in interest rates causes an increased supply of domestic currency to come on to the foreign exchange market; this causes the exchange rate to depreciate; this will cause increased exports (an injection) and reduced imports (a withdrawal) and hence a multiplied rise in national income.

11.6 MONEY SUPPLY, AGGREGATE DEMAND AND INFLATION

How will changes in money supply affect spending, output and prices?

There is considerable debate around the impact of changes in the money supply on output and inflation. The debate can be understood in terms of the transmission mechanisms that we looked at in the previous section, illustrated in Figure 11.12 (see page 335).

There are two key questions.

- First, to what extent will a change in money supply affect interest rates and how much, in turn, will a change in interest rates affect aggregate demand? In other words, what is the strength of the interest rate and exchange rate transmission mechanisms?
- Second, how much will a change in aggregate demand affect real output and how much will it merely result in higher prices?

The 1970s saw the rise of monetarism. The most famous advocate of monetarism was Milton Friedman, who argued that inflation can be attributed entirely to increases in the money supply. The faster money supply expands, the higher the rate of inflation will be. New classical economists of today take a similar view. Excessive expansion of the money supply will lead simply to inflation. According to monetarist and new classical economists, therefore, in answer to the first question, the mechanisms are strong: a rise in money supply directly affects aggregate demand. But any rise in aggregate demand will lead not to an increase in output (at least, according to monetarists, not in the long run) but merely to higher prices.

Keynesians, by contrast, see a much looser association between money and prices. The amount that a rise in money supply leads to higher aggregate demand and higher output depends on circumstances.

The debate can best be understood in terms of the **quantity theory of money**. The theory is simply that the level of prices in the economy depends on the quantity of money: the greater the supply of money, the higher the level of prices will be.

A development of the quantity theory is the *equation of exchange*. Focusing on this equation is the best way of understanding the debate over the relationship between money and prices.

The equation of exchange

The **equation of exchange** shows the relationship between the money value of spending and the money value of output (nominal GDP). This identity may be expressed as follows:

TC 14 p 249

$$MV = PY$$

M is the supply of money in the economy (e.g. M4). V is its **velocity of circulation**. This is the number of times per year that money is spent on buying goods and services that have been produced in the economy that year (real GDP). P is the level of prices of domestically produced goods and services,

Definitions

Quantity theory of money The price level (P) is directly related to the quantity of money in the economy (M).

Equation of exchange $MV = PY$. The total level of spending on GDP (MV) equals the total value of goods and services produced (PY) that go to make up GDP.

Velocity of circulation The number of times annually that money on average is spent on goods and services that make up GDP.

expressed as an index, where the index is 1 in a chosen base year (e.g. 2000). Thus, if prices today are double those in the base year, P is 2. Y is *real* national income (real GDP): in other words, the quantity of national output produced in that year measured in base-year prices.

PY is thus nominal GDP, i.e. GDP measured at current prices (see the Appendix to Chapter 9, page 270, and Web Appendix 9.1). For example, if GDP at base-year prices (Y) is £1 trillion and the price index is 2, then GDP at current prices (PY) is £2 trillion.

Pause for thought

If the money supply is cut by 10 per cent, what must happen to the velocity of circulation if there is no change in GDP at current prices?

MV is the total spending on the goods and services that make up GDP – in other words, (nominal) aggregate demand. For example, if money supply is £500 billion, and money, as it passes from one person to another, is spent on average four times a year on national output, then total spending (MV) is £2 trillion a year. But this too *must* equal GDP at current prices. The reason is that what is spent on output (by consumers, by firms on investment, by the government or by people abroad on exports) must equal the value of goods produced (PY).

The equation of exchange (or 'quantity equation') is true by definition. MV is *necessarily* equal to PY because of the way the terms are defined. Thus a rise in MV *must* be accompanied by a rise in PY.

What a change in M does to P, however, is a matter of debate. The controversy centres on the impact of changes in the money supply on aggregate demand and then on the impact of changes in aggregate demand on output. We have seen that the latter depends crucially on the nature of the aggregate supply. We develop this further in the next chapter.

We now focus now on the relationship between money supply and aggregate demand, beginning with the short-run relationship.

Money and aggregate demand

The short run

TC 3 p 10

According to the interest-rate and exchange-rate transmission mechanisms, the impact of an increase in the money supply can be summarised as follows:

1. A rise in money supply will lead to a fall in the rate of interest.
2. The fall in the rate of interest will lead to a rise in investment and other forms of borrowing. It will also lead to a fall in the exchange rate and hence a rise in exports and a fall in imports.
3. The rise in investment, and the rise in exports and fall in imports, will mean a rise in aggregate demand.

However, there is considerable debate over how these transmission mechanisms function.

How interest-rate elastic is money demand? The demand for money as a means of storing wealth (the assets motive) can be large and highly responsive to changes in interest rates on alternative assets. Indeed, large sums of money move around the money market as firms and financial institutions respond to and anticipate changes in interest rates. Therefore, the demand-for-money curve in Figure 11.9 (see page 332) could be relatively flat. This is important because, following an increase in money supply, only a relatively small fall in interest rates on bonds and other assets may be necessary to persuade people to hold all the extra money in bank accounts. This greatly slows down the average speed at which money circulates. The fall in V may virtually offset the rise in M.

TC 10 p 49

Indeed, this was one of the consequences of the large increases in money supply under the programmes of quantitative easing. These were adopted by central banks around the world in an attempt to stimulate recovery from the recession in 2009–11 that followed the financial crisis. But, although money supply rose, much of it was held in additional reserves rather than being spent.

The more sensitive the demand for money is to changes in the rate of interest, the less impact changes in money supply have on aggregate demand.

How stable is the money demand function? Another criticism is that the demand for money is unstable and so the demand-for-money curve in Figure 11.8 is frequently moving. People hold speculative balances of money when they anticipate that the prices of other assets, such as shares, bonds and bills, will fall (and hence the rate of return or interest on these assets will rise).

TC 11 p 61

There are many factors that could affect such expectations, such as changes in foreign interest rates, changes in exchange rates, statements of government intentions on economic policy, good or bad industrial news, or newly published figures on inflation or money supply. With an unstable demand for money, it is difficult to predict the effect of a change in money supply on interest rates and so on aggregate demand.

It has been largely for this reason that most central banks have preferred to control interest rates directly, rather than indirectly by controlling the money supply. We examine the conduct of monetary policy in Chapter 13.

TC 10 p 49

How interest-rate elastic is spending? The problem here is that investment may be insensitive to changes in interest rates. Businesses are more likely to be influenced in their decision to invest by predictions of the future buoyancy of markets. Interest rates do have *some* effect on businesses' investment decisions, but the effect is unpredictable, depending on the confidence of investors.

Where interest rates are likely to have a stronger effect on spending is via mortgages. If interest rates go up, and mortgage rates follow suit, people will suddenly be faced with higher monthly repayments (debt servicing costs) and will therefore have to cut down their expenditure on goods and services.

TC 10 p 49

How interest-rate sensitive is the exchange rate? Also the amount that the exchange rate will depreciate is uncertain, since exchange rate movements, as we shall see in Chapter 15, depend crucially on expectations about trade prospects and about future world interest rate movements. Thus the effects on imports and exports are also uncertain.

To summarise: the effects on total spending of a change in the money supply *might* be quite strong, but they could be weak. In other words, the effects are highly unpredictable.

$$M\uparrow \rightarrow V\downarrow (?) \rightarrow MV?$$

Keynesians use these arguments to criticise the use of monetary policy as a means of managing aggregate demand.

The long run

In the long run, there is a stronger link between money supply and aggregate demand. In fact, monetarists claim that in the long run V is determined *totally independently* of the money supply (M). Thus an increase in M will leave V unaffected and hence will directly increase expenditure (MV):

$$M\uparrow \rightarrow M\bar{V}\uparrow$$

where the bar over the V term means that it is **exogenously** determined: i.e. determined *independently* of M. But why do they claim this?

If money supply increases over the longer term, people will have more money than they require to hold. They will spend this surplus. Much of this spending will go on goods and services, thereby directly increasing nominal aggregate demand.

The theoretical underpinning for this is given by the *theory of portfolio balance*. People have a number of ways of holding their wealth. They can hold it as money, or as financial assets such as bills, bonds and shares, or as physical assets such as houses, cars and televisions. In other words, people hold a whole portfolio of assets of varying degrees of liquidity – from cash to central heating.

If money supply expands, people will find themselves holding more money than they require: their portfolios are 'unnecessarily liquid'. Some of this money will be used to purchase financial assets and some, possibly after a period of time, to purchase *goods and services*. As more assets are purchased, this will drive up their price. This will effectively reduce their 'yield'. For bonds and other *financial* assets, this means a reduction in their rate of interest. For goods and services, it means an increase in their price relative to their usefulness or 'utility'.

The process will stop when a balance has been restored in people's portfolios. In the meantime, there will have been extra consumption and hence an increase in aggregate demand.

Definition

Exogenous variable A variable whose value is determined independently of the model of which it is part.

Recap

1. The quantity equation $MV = PY$ can be used to analyse the possible relationships between money and prices.
2. In the short run, the velocity of circulation (V) may vary inversely, but unpredictably, with the money supply (M). The reason is that changes in money supply will have unpredictable and possibly rather weak effects on interest rates, and, similarly, changes in interest rates will have unpredictable and probably rather weak effects on aggregate demand. Thus spending (MV) will change by possibly only a small and rather unpredictable amount.
3. In the long run, there is a stronger link between money supply and aggregate demand.

QUESTIONS

1. Imagine that the banking system receives additional deposits of £100 million and that all the individual banks wish to retain their current liquidity ratio of 20 per cent.
   a. How much will banks choose to lend out initially?
   b. What will happen to banks' liabilities when the money that is lent out is spent and the recipients of it deposit it in their bank accounts?
   c. How much of these latest deposits will be lent out by the banks?
   d. By how much will total deposits (liabilities) eventually have risen, assuming that none of the additional liquidity is held outside the banking sector?
   e. How much of these are matched by (i) liquid assets; (ii) illiquid assets?
   f. What is the size of the bank deposits multiplier?
   g. If one half of any additional liquidity is held outside the banking sector, by how much less will deposits have risen compared with (d) above?
2. What are meant by the terms *narrow money* and *broad money*? Does broad money fulfil all the functions of money?
3. Do credit and/or debit cards count as (broad) money?
4. How does money aid the specialisation and division of labour?
5. What enables banks safely to engage in both maturity transformation and risk transformation?
6. Why do banks hold a range of assets of varying degrees of liquidity and profitability?
7. What is meant by the securitisation of assets? How might this be (a) beneficial and (b) harmful to banks and the economy?
8. What were the causes of the credit crunch and the banking crisis of the late 2000s?
9. Analyse the possible effects on banks' balance sheets of the following:
   a. the Basel III regulatory requirements.
   b. the UK bank levy.
10. What is measured by the CET1 ratio? What measures would a bank need to take to increase its CET1 ratio?
11. What is meant by macro prudential regulation? How has this concept influenced changes to the Basel regulatory framework?
12. Assume that banks choose to operate a 20 per cent liquidity ratio and receive extra cash deposits of £10 million. Assume also that the general public does not wish to hold a larger total amount of cash balances outside the banks:
    a. How much credit will ultimately be created?
    b. By how much will total deposits have expanded?
    c. What is the size of the bank deposits multiplier?
13. If the government reduces the size of its public-sector net cash requirement, why might the money supply nevertheless increase more rapidly?
14. What is mean by a credit cycle? What impact do such cycles have on monetary growth?
15. Why might the relationship between the demand for money and the rate of interest be an unstable one?
16. What effects will the following have on the equilibrium rate of interest? (You should consider which way the demand and/or supply curves of money shift.)
    a. Banks find that they have a higher liquidity ratio than they need.
    b. A rise in incomes.
    c. A growing belief that interest rates will rise from their current level.
17. Trace through the effect of a fall in the supply of money on aggregate demand. What will determine the size of the effect?
18. If V is constant, will (a) a £10 million rise in M give a £10 million rise in MV and (b) a 10 per cent rise in M give a 10 per cent rise in MV? (Test your answer by fitting some numbers to the terms.)
19. If both V and Y are constant, will (a) a £10 million rise in M lead to a £10 million rise in P and (b) a 10 per cent rise in M lead to a 10 per cent rise in P? (Again, try fitting some numbers to the terms.)
20. What would be the implications of different assumptions about the V and Y terms in the quantity equation $MV = PY$ for the effectiveness of monetary policy to control inflation?

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

11.1 **Barter: its use in Russia in the 1990s.** When barter was used as an alternative to money.

11.2 **The attributes of money.** What makes something, such as metal, paper or electronic records, suitable as money?

11.3 **From coins to bank deposit money.** This case traces the evolution of modern money.

11.4 **Changes in the banking industry.** This considers the changing face of the banking industry and asks whether bigger is better.

11.5 **Are the days of cash numbered?** Does the increased use of credit and debit cards and direct debits mean that cash will become obsolete?

11.6 **Gresham's law.** Historic examples of 'bad' money driving out 'good' money as people hoard the 'good' money.

11.7 **German banking.** This case compares the tradition of German banks with that of UK retail banks. Although the banks have become more similar in recent years, German banks have a much closer relationship with industry.

11.8 **Bailing out the banks.** An overview of the concerted efforts made to rescue the banking system in the crisis of 2007–9.

11.9 **Parallel money markets.** A description of the variety of short-term financial instruments available in the parallel money markets.

11.10 **UK monetary aggregates.** This case shows how money supply is measured using both UK measures and eurozone measures.

11.11 **Making money grow.** A light-hearted illustration of the process of credit creation.

11.12 **Consolidated MFI balance sheet.** A look at the consolidated balance sheet of UK monetary financial institutions (banks, building societies and the Bank of England).

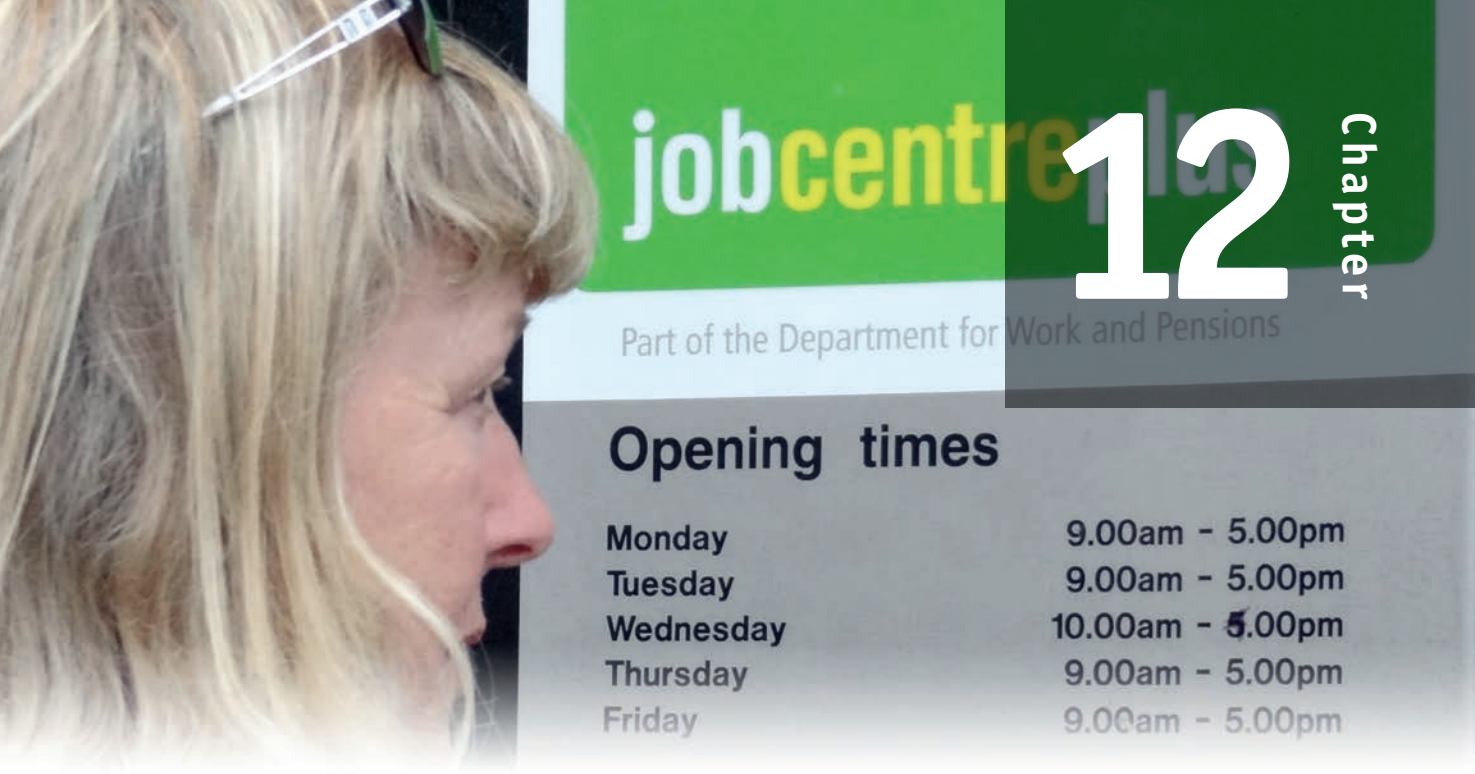
11.13 **The equation of exchange.** This examines two more versions that are commonly used: the Fisher version and the Cambridge version.

WEB APPENDICES

11.1 **The money multiplier.** This appendix shows how the money multiplier is calculated.

11.2 ***IS/LM* analysis.** This appendix develops a model that brings together the goods and money markets in one diagram.

11.3 ***IS/MP* analysis.** This develops a more recent model of the goods and money markets based on the assumption that the central bank targets the rate of inflation.



Chapter 12

Output, unemployment and inflation

In the previous chapters we have considered some of the influences on aggregate demand and aggregate supply as well as the importance of money and the financial system. In this chapter we draw on this analysis to consider the relationship between output, unemployment and inflation.

We start by looking at unemployment: its nature and causes. We see that it is determined partly by aggregate demand and supply, but this time for labour rather than for goods and services. We also look at other causes.

We have seen that the economy's price level is determined by aggregate demand and supply. In this chapter we will consider what determines the rate of *increase* in prices. In other words, what determines the rate of inflation?

Having looked at inflation and unemployment separately, we then see how they are related. To do this, we apply first the *AD*/AS model before then considering the Phillips curve – a curve that plots inflation against unemployment. A crucial element in this relationship is people's expectations of inflation. This is the topic for Section 12.4 when we look at the 'expectations-augmented Phillips curve (*EAPC*)'.

In recent years, many governments around the world have delegated the operation of monetary policy to central banks and set targets for the rate of inflation. In Section 12.5 we consider why governments would choose to do so. We analyse the profound effect this has had on people's expectations and on the relationships between output, employment and inflation.

After studying this chapter, you should be able to answer the following questions:

- What are the causes of unemployment?
- What factors influence long-term patterns in rates of unemployment?
- What are the causes of inflation?
- What factors affect the relationship between output, unemployment and inflation?
- How can we incorporate expectations into the Phillips curve?
- Why would governments choose to delegate monetary policy to central banks and adopt inflation rate targets?
- How does a policy of targeting a rate of inflation affect the relationship between output, unemployment and inflation?

12.1 UNEMPLOYMENT

If people want to consume more goods, why are so many people out of work?

Measuring unemployment

As we saw in Chapter 9, unemployment can be expressed either as a number (e.g. 2 million) or, more commonly, as a percentage (e.g. 5 per cent). When expressed as a percentage, it is as a percentage of the total labour force, i.e. those in employment plus those unemployed. We are then able to talk about unemployment rates.

To be able to compare unemployment rates both over time and between countries, we use a **standardised unemployment rate**. Sometimes known as ILO unemployment, this is the measure used by the International Labour Organisation (ILO) and the Organisation for Economic Co-operation and Development (OECD), two international organisations that publish unemployment statistics for many countries.

In this measure, the unemployed are defined as people of working age who are without work, available to start work and *actively seeking employment* or waiting to take up an appointment. The figures are compiled from the results of national labour force surveys. In the UK the labour force survey is conducted quarterly.

Countries also compile unemployment statistics using administrative measures, typically based on unemployment-related benefits. This is known as **claimant unemployment**. The claimant count is simply a measure of all those in receipt of unemployment-related benefits. In the UK, claimants receive the Jobseeker's Allowance or Universal Credit.

Claimant statistics have the advantage of being very easy to collect. However, they exclude all those of working age available for work at current wage rates, but who are *not* eligible for benefits. If the government changes the eligibility conditions so that fewer people are now eligible, this will reduce the number of claimants and hence the official number unemployed, even if there has been no change in the numbers with or without work.

Changing eligibility conditions make historical comparisons using claimant statistics more difficult. Furthermore, different countries adopt different approaches to unemployment benefit. In turn, this makes country comparisons more difficult. Recognising these difficulties, the standardised unemployment rate has become the generally accepted measure of unemployment.

Unemployment trends

Cyclical movements

Economies are inherently volatile and the fluctuations in economic activity give rise to the business cycle. Hence, unemployment fluctuates with the business cycle. In recessions, such as those experienced by most countries in the early 1980s, early 1990s and the late 2000s, unemployment tends to rise. In boom years, such as the late 1980s, late 1990s and mid-2000s, it tends to fall. Figure 12.1 shows the cyclical movements in unemployment for selected countries

TC 3 p 10

Longer-term movements

As well as experiencing fluctuations in unemployment, most countries have experienced long-term changes in average unemployment rates. This is illustrated in Table 12.1, which shows average unemployment rates across a sample of countries since the 1960s.

It shows how unemployment rates rose across the 1970s and the 1980s. Rates then fell through the late 1990s and early 2000s in some countries but remained stubbornly high in others. The global financial crisis then caused unemployment rates to rise generally, before falling in some countries, but again with some experiencing a persistence of unemployment. Lockdown measures during the pandemic then severely depressed economic activity. The impact on unemployment was partially mitigated in some countries through job retention schemes. However, concerns remained about the longer-term effects from structural changes in the labour market arising from or exacerbated by the pandemic.

Composition of unemployment

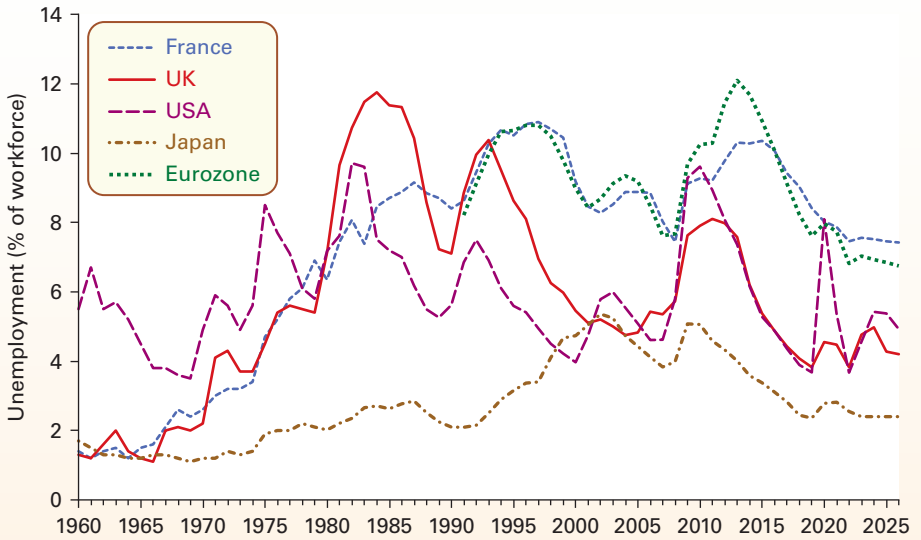
In many countries, female unemployment has traditionally been higher than male unemployment. Causes have included differences in education and

Definitions

Standardised unemployment rate The measure of the unemployment rate used by the ILO and OECD. The unemployed are defined as persons of working age who are without work, available for work and actively seeking employment.

Claimant unemployment Those in receipt of unemployment-related benefits.

Figure 12.1 Standardised unemployment rates



Notes: Data from 2022 based on forecasts; Eurozone = 19 countries using the euro (as of October 2022)

Source: Based on data in *AMECO Database* (European Commission) to 1980 and *World Economic Outlook* (IMF, October 2022)

training, discrimination by employers, more casual or seasonally related employment among women, and other social factors. In many countries, as highlighted in Table 12.2, the position has changed in recent years. An important reason has been the decline in many of the older industries, such as coal and steel, which employed mainly men. Across the whole EU, male and female unemployment rates are virtually the same.

Table 12.2 does, however, show some stark differences across different age groups. Unemployment rates in the under-25 age group are higher than the average, and substantially so in many countries. Higher youth unemployment rates can be explained by the suitability (or unsuitability) of the qualifications of school leavers, the attitudes of employers to young people, and the greater willingness of young people to spend time unemployed looking for a better job or waiting to start a further or higher education course. The difference in rates is less in Germany, which has a well-established apprenticeship system.

Table 12.1 Average unemployment by decade (%)

	1960s	1970s	1980s	1990s	2000s	2010s	1960–2023
Australia	1.8	3.9	7.6	8.8	5.5	5.5	5.5
Canada	5.0	6.7	9.4	9.6	7.0	6.9	7.5
Denmark	1.3	3.3	7.0	7.3	4.7	6.6	5.0
France	1.7	4.4	8.6	9.7	8.4	9.6	7.1
Germany	0.7	2.0	5.8	7.4	8.5	4.5	4.7
Ireland	5.4	7.5	14.2	12.1	5.7	10.7	9.0
Italy	4.9	6.0	8.5	10.7	8.2	10.9	8.3
Japan	1.3	1.7	2.5	3.1	4.7	3.6	2.8
Netherlands	1.1	4.7	10.3	7.2	4.8	6.6	5.7
Spain	2.5	4.5	16.4	19.5	11.4	20.5	12.6
UK	1.6	4.4	9.9	8.2	5.4	6.0	5.8
USA	4.8	6.2	7.3	5.8	5.5	6.3	5.9

Notes: Data for 2022 and 2023 based on forecasts; figures for Germany based on West Germany only up to 1991

Source: *AMECO Database* (European Commission, Economic and Financial Affairs, May 2022)

BOX 12.1 THE COSTS OF UNEMPLOYMENT EXPLORING ECONOMICS

Is it just the unemployed who suffer?

The most obvious cost of unemployment is to the *unemployed themselves.* There is the direct financial cost of the loss in their earnings, measured as the difference between their previous wage and their unemployment benefit. Then there are the personal costs of being unemployed. The longer people are unemployed, the more dispirited they may become. Their self-esteem is likely to fall, and they are more likely to succumb to stress-related illness.

Beyond the unemployed themselves, there are the costs to the *family and friends* of the unemployed. Personal relations can become strained, and there may be an increase in domestic violence and the number of families splitting up.

Then there are the *broader costs to the economy.* Unemployment benefits are a cost borne by taxpayers. There may also have to be extra public spending on benefit offices, social services, health care and the police. What is more, unemployment represents a loss of output. In other words, actual output is below potential output. Apart from the lack of income to the unemployed themselves, this under-utilisation of resources leads to lower incomes for other people too:

- The government loses tax revenues, since the unemployed pay no income tax or national insurance, and, given that the unemployed spend less, they pay less VAT and excise duties.
- Firms lose the profits that could have been made, had there been full employment.
- Other workers lose any additional wages they could have earned from the higher national output.

What is more, the longer people remain unemployed, the more deskilled they tend to become. This scarring effect reduces potential as well as actual income.

Finally, there is some evidence that higher unemployment leads to increased crime and vandalism. This obviously imposes a cost on the sufferers.

The costs of unemployment are to some extent offset by benefits. If workers voluntarily quit their jobs to look for a better one, then they must reckon that the benefits of a better job more than compensate for their temporary loss of income. From the nation's point of view, a workforce that is prepared to quit jobs and spend a short time unemployed will be a more adaptable, more mobile workforce – one that is responsive to changing economic circumstances. Such a workforce will lead to greater allocative efficiency in the short run and more rapid economic growth over the longer run.

Long-term involuntary unemployment is quite another matter. The costs clearly outweigh any benefits, both for the individuals concerned and for the economy as a whole. A demotivated, deskilled pool of long-term unemployed is a serious economic and social problem.

How might an economist set about measuring the various costs of unemployment to family, friends and society at large?

Conduct a literature search on the concept of unemployment scarring (also known as labour market scarring). Briefly summarise your findings in a short paper aimed at a newly appointed government minister with responsibility for employment.

Unemployment and the labour market

TC4 p13

We now turn to the causes of unemployment. These causes fall into two broad categories: *equilibrium* unemployment and *disequilibrium* unemployment. To make clear the distinction between the two, it is necessary to look at how the labour market works.

Figure 12.2 shows the aggregate demand for labour and the aggregate supply of labour: that is, the total demand and supply of labour in the whole economy. The *real* average wage rate is plotted on the vertical axis. This is the average wage rate expressed in terms of its purchasing power: in other words, after taking prices into account. In other words, the real wage is the nominal wage corrected for inflation.

TC5 p17

The **aggregate supply of labour curve** (AS_L) shows the number of workers *willing to accept jobs* at each real wage rate. This curve is relatively inelastic, since the size of the workforce at any one time cannot change significantly. Nevertheless, it is not totally inelastic because (a) a higher wage rate will encourage some people to enter the labour market (e.g. parents raising children), and (b) the unemployed will be more willing to accept job offers rather than continuing to search for a better paid job.

The **aggregate demand for labour curve** (AD_L) slopes downwards. The higher the real wage rate, the more firms will attempt to economise on labour and to substitute other factors of production for labour.

The labour market is in equilibrium at a wage of W_e, where the demand for labour equals the supply.

Definitions

Aggregate supply of labour curve A curve showing the total number of people willing and able to work at different average real wage rates.

Aggregate demand for labour curve A curve showing the total demand for labour in the economy at different average real wage rates.

Table 12.2 Standardised unemployment rates by age and gender, average 2010–21

	Less than 25 years			25 to 74 years			All ages		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
EU – 27 countries (from 2020)	20.6	20.5	20.6	7.8	8.3	8.0	9.0	9.4	9.2
Eurozone – 19 countries (from 2015)	20.9	20.6	20.8	8.4	9.0	8.7	9.6	10.0	9.8
Austria	11.0	10.0	10.5	5.2	4.7	5.0	6.0	5.3	5.6
Belgium	20.5	18.5	19.6	6.4	6.1	6.3	7.6	7.0	7.3
Denmark	14.2	11.9	13.1	5.0	5.5	5.2	6.3	6.5	6.4
France	22.3	25.6	23.8	7.9	7.6	7.8	9.4	9.3	9.3
Germany	8.6	7.1	7.9	4.2	3.6	3.9	4.7	4.0	4.4
Greece	40.7	50.5	45.2	16.3	23.3	19.3	17.7	25.1	20.9
Ireland	23.6	17.1	20.6	9.1	7.5	8.4	10.9	8.8	9.9
Italy	32.5	36.3	34.1	8.3	10.1	9.1	9.9	11.7	10.6
Netherlands	12.5	10.4	11.4	4.6	5.8	5.1	5.9	6.6	6.2
Poland	18.1	20.0	18.9	5.5	6.1	5.8	6.6	7.2	6.9
Portugal	26.8	29.8	28.2	9.7	10.1	9.9	11.1	11.5	11.3
Spain	43.7	43.0	43.4	16.6	19.2	17.8	18.5	20.9	19.6
Sweden	22.0	20.3	21.2	6.0	5.9	5.9	7.9	7.7	7.8

Source: Based on data from *Employment and Unemployment (LFS)* (Eurostat, European Commission, 2022)

Figure 12.2 Disequilibrium unemployment

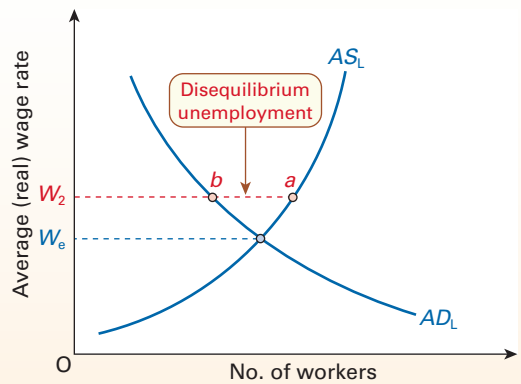
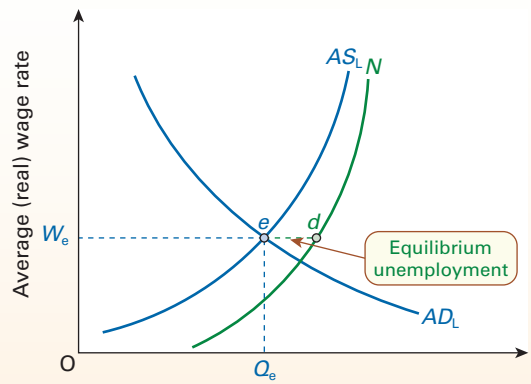


Figure 12.3 Equilibrium unemployment



If the wage rate were above W_e, the labour market would be in a state of disequilibrium. At a wage rate of W_2, there is an excess supply of labour of $a - b$. This is called **disequilibrium unemployment**.

For disequilibrium unemployment to occur, two conditions must hold:

- The aggregate supply of labour must exceed the aggregate demand.
- There must be a 'stickiness' in wages. In other words, the wage rate must not immediately fall to W_e.

Definition

Disequilibrium unemployment Unemployment resulting from real wages in the economy being above the equilibrium level.

Even when the labour market *is* in equilibrium, however, not everyone looking for work will be employed. Some people will hold out, hoping to find a better job. The curve N in Figure 12.3 shows the total number in the labour force. The horizontal difference between it and the aggregate supply of labour curve (AS_L) represents the excess of people looking for work over those actually willing to accept jobs. Q_e represents the equilibrium level of employment and the distance $d - e$ represents the **equilibrium level of unemployment**. This is sometimes known as the **natural level of unemployment**.

TC 8 p 37

Types of disequilibrium unemployment

There are three possible causes of disequilibrium unemployment.

Real-wage unemployment

Real-wage unemployment is disequilibrium unemployment caused by real wages being driven up above the market-clearing level. In Figure 12.2, the wage rate is driven up above W_e. Some argue that it is the result of trade unions using their monopoly power to drive wages above the market-clearing level or preventing the unemployed from competing wages down.

Even though unions have the power to drive up wages in some industries, their power to do so has waned in recent years. Labour markets have become more flexible (see Case Study 7.2 on the student website). What is more, the process of globalisation has meant that many firms face intense competition from rivals in China, India and many other countries. This makes it impossible for them to concede large pay increases. In many cases, they can simply use labour in other countries if domestic labour is too expensive. For example, many firms employ call-centre workers in India, where wages are much lower.

As far as the national minimum wage is concerned, evidence from the UK suggests that the rate has not been high enough to have significant adverse effects on employment (see Box 7.4 on page 182).

Demand-deficient or cyclical unemployment

Demand-deficient or cyclical unemployment is associated with recessions. As the economy moves into recession, consumer demand falls. Firms find that they are unable to sell their current level of output. For a time, they may be prepared to build up stocks of unsold goods, but sooner or later they will start to cut back on production and cut back on the amount of labour they employ. In Figure 12.2 the AD_L curve shifts to the left. The deeper the recession becomes and the longer it lasts, the higher will demand-deficient unemployment become. Only later, as the economy recovers and begins to grow again, will demand-deficient unemployment start to fall again.

TC 3 p 10

But if there is surplus labour as a result of a leftward shift of AD_L, why do real wages not fall so as to eliminate the surplus? There are two main reasons for this.

Efficiency wages. The argument here is that wage rates fulfil two functions. The first is the traditional one of balancing the demand and supply of labour. The second is that of motivating workers. If real wage rates are reduced when there is a surplus of labour, then those workers already in employment may become dispirited and work less hard. If, on the other hand, firms keep wage rates up, then by maintaining a well-motivated workforce, by cutting down on labour turnover and by finding it easier to attract well-qualified labour, firms may find their costs are reduced: a higher real wage is thus more profitable for them.

TC 6 p 24

The maximum-profit real wage rate – the **efficiency wage rate** – is likely to be above the market-clearing real wage rate. Demand-deficient unemployment is likely to persist – more people would like to work at the efficiency wage than there are jobs available.

Insider power. If those still in employment (the insiders) are members of unions, while those out of work (the outsiders) are not, or if the insiders have special skills or knowledge that give them bargaining power with employers, while the outsiders have no influence, then there is no mechanism whereby the surplus labour – the outsiders – can drive down the real wage rate and eliminate the demand-deficient unemployment.

KI 19 p 122

Even if real wages *did* fall, it may not solve the problem of demand-deficient unemployment. In fact, it might even make the problem worse. The reason is that this general cut in real wages throughout the economy would reduce workers' real incomes and

Definitions

Equilibrium ('natural') unemployment The difference between those who would like employment at the current wage rate and those willing and able to take a job.

Real-wage unemployment Disequilibrium unemployment caused by real wages being driven up above the market-clearing level.

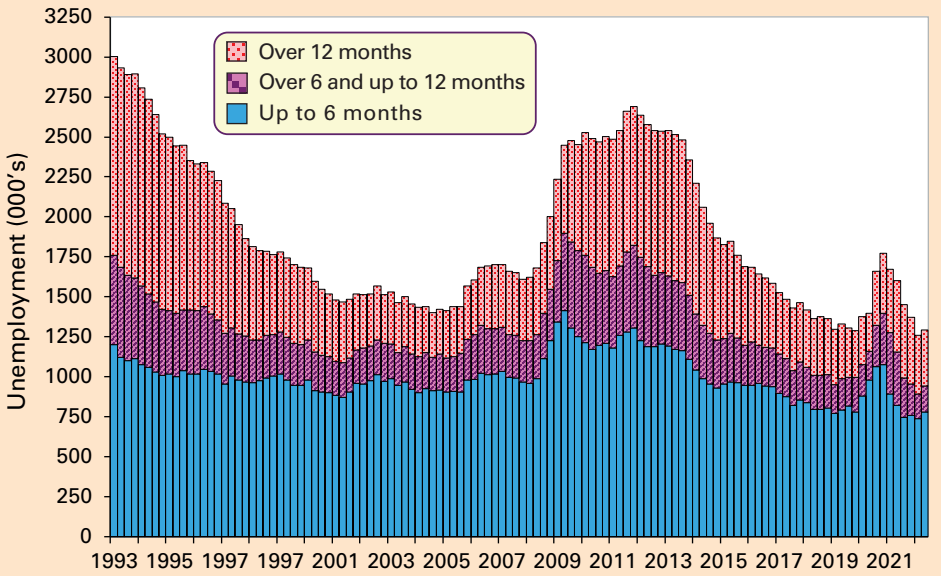
Demand-deficient or cyclical unemployment Disequilibrium unemployment caused by a fall in aggregate demand with no corresponding fall in the real wage rate.

Efficiency wage rate The profit-maximising wage rate for the firm after taking into account the effects of wage rates on worker motivation, turnover and recruitment.

BOX 12.2 **THE DURATION OF UNEMPLOYMENT** CASE STUDIES & APPLICATIONS

Taking a dip in the unemployment pool

UK standardised unemployment by duration



Note: Figures relate to people aged 16 and over

Source: Based on series YBWF, YBWG and YBWH (Office for National Statistics)

A few of the unemployed may never have had a job and maybe never will. For most, however, unemployment lasts only a certain period. For some it may be just a few days while they are between jobs. For others it may be a few months. For others – the long-term unemployed – it could be several years.

Long-term unemployment is normally defined as those who have been unemployed for over 12 months. The chart shows the composition of standardised unemployment in the UK by duration since the early 1990s. It shows how long-term unemployment fell from the mid-1990s until the global financial crisis in the late 2000s. As a result, the percentage classified as long-term unemployed, which had hit 45 per cent during 1994, fell to below 20 per cent in 2004, before rising to 37 per cent in 2014. It then fell back to 23 per cent in 2019, before rising following the pandemic to exceed 30 per cent in 2021.

But what determines the average duration of unemployment? There are three important factors here.

The number unemployed (the size of the stock of unemployment)

Unemployment is a 'stock' concept: it measures a quantity of people unemployed at a particular point in time. The higher the stock of unemployment, the longer will tend to be the duration of unemployment. There will be more people competing for vacant jobs.

KI 23 p 173

The rate of inflow and outflow from the stock of unemployment

The people making up the unemployment total are constantly changing. Each week some people are made redundant or quit their jobs. They represent an inflow to the stock of unemployment. Other people find jobs and thus represent an outflow from the stock of unemployment. Unemployment is often referred to as 'the pool of unemployment'.

If the inflow of people into the unemployment pool exceeds the outflow, the pool of unemployed people will rise. The duration of unemployment will depend on the rate of inflow and outflow.

The bigger the flows are relative to the total number unemployed, the less will be the average duration of unemployment. This is because people move into and out of the unemployment pool more quickly, and hence their average stay will be shorter.

The phase of the business cycle

The duration of unemployment will also depend on the phase of the business cycle. At the onset of a recession, unemployment will rise, but as yet the average length of unemployment is likely to have been relatively short. Once a recession has lasted for a period of time, however, people will on average have been out of work longer; and this long-term unemployment is likely to persist even when the economy is pulling out of recession.

TC 3 p 10

1. *If the number unemployed exceeded the total annual outflow, what could we conclude about the average duration of unemployment?*
2. *Make a list of the various inflows to and outflows from employment from and to (a) unemployment; (b) outside the workforce.*

From the Office for National Statistics, download series YBWH (unemployed over 12 months) and MGSC (total unemployed) and calculate the percentage of the total unemployed who have unemployed for over 12 months. Plot this in a line chart before writing a short summary of your findings.

hence reduce their *consumption of goods*. As the aggregate demand for goods fell, there would be a further reduction in demand for labour: the aggregate demand for labour curve would shift further to the left. The fall in real wages would be 'chasing' a leftward-shifting AD_L curve. As the real wage fell towards equilibrium, so the equilibrium real wage itself would be falling and equilibrium would not be reached. There would still be demand-deficient unemployment.

Pause for thought

If this analysis is correct, namely that a reduction in real wages will reduce the aggregate demand for goods, what assumption must we make about the relative proportions of wages and profits that are spent (given that a reduction in real wage rates will lead to a corresponding increase in rates of profit)?

Because demand-deficient unemployment fluctuates with the business cycle, it is frequently referred to as 'cyclical unemployment'. It is also known as 'Keynesian unemployment', after John Maynard Keynes who saw a deficiency of aggregate demand as the cause of the high unemployment between the two world wars. The more that aggregate demand fluctuates and the greater the stickiness in wages, the more significant the cyclical component of unemployment becomes.

Growth in the labour supply

If labour supply rises with no corresponding increase in the demand for labour, the equilibrium real wage rate will fall. If the real wage rate is 'sticky' downwards, unemployment will occur. This tends not to be such a serious cause of unemployment as demand deficiency, since the supply of labour changes relatively slowly. Nevertheless, there is a problem of providing jobs for school and university leavers each year with the sudden influx of new workers on to the labour market.

Types of equilibrium unemployment (or natural unemployment)

If you look at Table 12.1 on page 343, you can see how unemployment in many countries was higher in the 1980s and 1990s than in the previous two decades. Part of the reason for this was the growth in equilibrium unemployment. In the 2000s, unemployment fell in many countries – at least until the financial crisis of 2007/8. Again, part of the reason for this was a change in equilibrium unemployment, but this time a fall.

Although there may be overall *macro*economic equilibrium, with the *aggregate* demand for labour equal to the *aggregate* supply, and thus no disequilibrium unemployment, at a *micro*economic level supply and demand may not match. In other words, there may be vacancies in some parts of the economy, but an excess of labour (unemployment) in others. This is equilibrium unemployment. There are various types of equilibrium unemployment.

Frictional (search) unemployment

Frictional (search) unemployment occurs when people leave their jobs, either voluntarily or because they are sacked or made redundant, and are then unemployed for a period of time while they are looking for a new job. They may not get the first job they apply for, despite a vacancy existing. The employer may continue searching, hoping to find a better-qualified person. Likewise, unemployed people may choose not to take the first job they are offered. Instead, they may continue searching, hoping that a better one will turn up.

The problem is that information is imperfect. Employers are not fully informed about what labour is available; workers are not fully informed about what jobs are available and what they entail. Both employers and workers, therefore, have to search: employers searching for the right labour and workers searching for the right jobs.

One obvious remedy for frictional unemployment is for there to be better job information through government job centres, private employment agencies or the media. Another much more controversial remedy is for the government to reduce the level of unemployment benefit. This will make the unemployed more desperate to get a job and thus prepared to accept a lower wage.

Structural unemployment

Structural unemployment occurs where the structure of the economy changes. Many countries have witnessed rapid structural change over recent years. Employment in some industries may expand while in others it contracts. There are two main reasons for this.

Definitions

Frictional (search) unemployment Unemployment that occurs as a result of imperfect information in the labour market. It often takes time for workers to find jobs (even though there *are* vacancies) and in the meantime they are unemployed.

Structural unemployment Unemployment that arises from changes in the pattern of demand or supply in the economy. People made redundant in one part of the economy cannot immediately take up jobs in other parts (even though there are vacancies).

A change in the pattern of demand. Some industries experience declining demand. This may be due to a change in consumer tastes. Certain goods may go out of fashion. Or it may be due to competition from other industries. For example, consumer demand may shift away from coal and to other fuels. This will lead to structural unemployment in mining areas.

A change in the methods of production (technological unemployment). New techniques of production often allow the same level of output to be produced with fewer workers. This is known as 'labour-saving technical progress'. Unless output expands sufficiently to absorb the surplus labour, people will be made redundant. This creates **technological unemployment**. Examples include the loss of jobs in the retail sector from the rise of e-retail and in banking from the increased use of electronic payments and online banking services. These trends have been accentuated by the COVID-19 pandemic.

Structural unemployment is sometimes referred to as 'mismatch unemployment'. Structural changes to an economy will tend to lead to a mismatch between the skills of the workforce and the needs of employers. There are often significant geographical mismatches with structural unemployment focused in particular regions of the country. When it does, it is referred to as **regional unemployment**. This is most likely to occur when particular industries are concentrated in particular areas. For example, the decline in the South Wales coal-mining industry led to high unemployment in the Welsh valleys.

The level of structural unemployment depends on three factors.

- The degree of *regional concentration* of industry. The more that industries are concentrated in particular regions, the greater the level of structural unemployment if particular industries decline will be.
- The *speed of change* of demand and supply in the economy. The more rapid the rate of technological change or the shift in consumer tastes, the more rapid will be the rate of redundancies.
- The *immobility of labour.* Where workers are less able or less willing to move to a new job, the higher will be the level of structural unemployment.

Seasonal unemployment

Seasonal unemployment occurs when the demand for certain types of labour fluctuates with the seasons of the year. This problem is particularly severe in holiday areas, such as Cornwall, where unemployment can reach very high levels in the winter months.

Long-term changes in unemployment

We are now able to analyse the impact of the equilibrium unemployment rate on actual unemployment rates. Consider the substantial rise in unemployment seen in the UK (and other developed economies) from the early 1970s to the mid 1980s and identified in Table 12.1. What could have led to the equilibrium rate rising?

The 1970s and 1980s were a period of rapid industrial change. The changes included the following:

- Dramatic changes in technology. The microchip revolution, for example, had led to many traditional jobs becoming obsolete.
- Competition from abroad. The introduction of new products from abroad, often of superior quality to domestic goods, or produced at lower costs, had led to the decline of many older industries: e.g. the textile industry.
- Shifts in demand away from the products of older labour-intensive industries to new 'high-tech' capital-intensive products.

The free market simply could not cope with these changes without a large rise in structural/technological unemployment. Labour was not sufficiently mobile – either geographically or occupationally – to move to industries where there were labour shortages or into jobs where there are skill shortages. A particular problem here was the lack of investment in education and training, with the result that the labour force was not sufficiently flexible to respond to changes in demand for labour.

From the mid 1980s, however, there were increasing signs that the UK labour market was becoming more flexible. People seemed more willing to accept that they would have to move from job to job throughout their career. At the same time, policies were introduced to improve training (see Section 13.4). The equilibrium rate was falling.

Another explanation for first the rise of equilibrium unemployment and later the fall is the phenomenon of 'hysteresis'.

Definitions

Technological unemployment Structural unemployment that occurs as a result of the introduction of labour-saving technology.

Regional unemployment Structural unemployment occurring in specific regions of the country.

Seasonal unemployment Unemployment associated with industries or regions where the demand for labour is lower at certain times of the year.

Hysteresis

KI 30 p 203

If a recession causes a rise in unemployment which is not then fully reversed when the economy recovers, there is a problem of **hysteresis**. This term, used in physics, refers to the lagging or persistence of an effect, even when the initial cause has been removed. In our context it refers to the persistence of unemployment even when the initial demand deficiency no longer exists.

The recessions of the early 1980s and early 1990s created a growing number of long-term unemployed who were both deskilled and demotivated. What is more, many firms, in an attempt to cut costs, cut down on training programmes. In these circumstances, a subsequent rise in aggregate demand would not simply enable the long-term unemployed to be employed again.

The recessions also caused a lack of investment and a reduction in firms' capacity. When demand recovered, many firms were unable to increase output and instead raised prices. Unemployment thus fell only modestly and inflation rose. The equilibrium rate had increased.

After 1992, however, the UK, along with most other countries, achieved a protracted period of sustained expansion, with no recession. Equilibrium unemployment began to fall. In other words, the hysteresis was not permanent. As firms increased their investment, the capital stock expanded; firms engaged in more training; the number of long-term unemployed fell.

Global economic shocks

The new millennium saw the global economy experience two 'once-in-a generation' shocks in relatively quick succession. Unemployment rates rose sharply in many countries following the financial crisis of the late 2000s.

However, in the UK the fall from a peak of 8.5 per cent in 2011 to below 5 per cent by 2016 led some to argue that hysteresis was now less of a problem. The principal reason, they argued, was the impact of greater labour market flexibility which allowed firms to retain workers by introducing part-time working or negotiating nominal wage freezes or cuts, thereby causing real wages to fall.

Also, increasing numbers of people were on 'zero-hour contracts'. This means that workers have no set number of hours per week and hours can be decreased (or increased) according to demand. Thus in a recession, employers can simply cut the number of hours offered to workers on such contracts.

To some extent, therefore, the problem of unemployment has been replaced by a problem of **underemployment** (see Section 9.1). One way in which this problem occurs is when people would like to work more hours than they are able to obtain, either in their current job or in an alternative job or in an additional part-time job. To be officially classified as underemployed, they must also be ready and available to work additional hours in a specified period and be currently working less than a specified threshold number of hours.

National measures of underemployment can vary significantly. The time-related underemployment rate in the UK is based on a threshold of 40 hours or less for people aged under 18, or 48 hours or less for people aged 18 and over. Individuals must be available to start work within two weeks.

Figure 12.4 shows time-related underemployment rates for the UK since 2002. Underemployment rates rose following the financial crisis from around 7 per cent in early 2008 to in excess of 10 per cent for much of the first half of the 2010s.

Rates then fell back until the outbreak of the COVID-19 pandemic and the restrictive measures put in place by governments around the world. This saw the underemployment rate rise to 9 per cent in late 2020, before labour shortages in some sectors, such as hospitality and leisure, agriculture, and transport, helped to drive down underemployment rates.

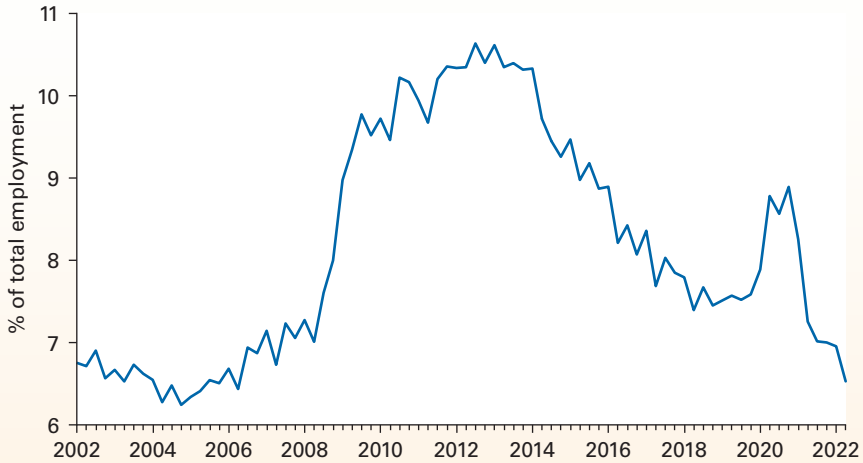
Meanwhile, unemployment rates in the first half of the 2010s remained significantly above pre-financial crisis levels in many countries in the eurozone, such as Greece, Portugal and Spain, which had to seek bailouts because of their high levels of debt. A condition of being granted bailouts was to reduce public-sector debt. This ruled out Keynesian expansionary fiscal policy. The very high levels of unemployment in these countries, especially amongst the young (rates of over 50 per cent in Greece and Spain in the 15–24 age group), resulted in a problem of entrenchment and hysteresis that made reductions in unemployment slow and difficult to achieve.

Definitions

Hysteresis The persistence of an effect even when the initial cause has ceased to operate. In economics it refers to the persistence of unemployment even when the demand deficiency that caused it no longer exists.

Underemployment International Labour Organisation (ILO) definition: a situation where people currently working less than 'full time' would like to work more hours (at current wage rates), either by working more hours in their current job, or by switching to an alternative job with more hours or by taking on an additional part-time job or any combination of the three.

Figure 12.4 UK rate of underemployment



Note: Rate of underemployment is not seasonally adjusted and is based on the proportion of people in total employment

Source: Based on data in *EMP16: Underemployment and overemployment* (ONS, August 2022)

Even in the UK, the problem of hysteresis remains relevant. First, despite the falling aggregate unemployment rate in the UK during the 2010s, long-term unemployment rates remained stubbornly high (see Box 12.2). Second, youth unemployment rates continued to be more than double the aggregate level (see Table 12.2). Third, the COVID-19 pandemic exacerbated long-term and youth unemployment rates. It also hastened the speed of structural change in the economy and therefore the labour market.

Recap

1. Unemployment can be measured as a number or as a percentage. When measured as a percentage, it is the percentage of the labour force (i.e. the employed plus the unemployed).
2. The two most common measures of unemployment are claimant unemployment (those claiming unemployment-related benefits) and ILO/OECD-defined standardised unemployment (those available for work and actively seeking work or waiting to take up an appointment). Historical and international comparisons are typically made using standardised unemployment rates.
3. Unemployment can be divided into disequilibrium and equilibrium unemployment.
4. Disequilibrium unemployment occurs when the average real wage rate is above the level that will equate the aggregate demand and supply of labour. It can be caused by unions or government pushing up wages (real-wage unemployment), by a fall in aggregate demand (demand-deficient unemployment), or by an increase in the supply of labour.
5. Equilibrium unemployment occurs when there are people unable or unwilling to fill job vacancies. This may be due to poor information in the labour market and hence a time lag before people find suitable jobs (frictional unemployment), to a changing pattern of demand or supply in the economy and hence a mismatching of labour with jobs (structural unemployment – specific types being technological and regional unemployment), or to seasonal fluctuations in the demand for labour.
6. Data for the UK show a growth in equilibrium unemployment in the 1970s and 1980s. Reasons include rapid technological changes and a persistence of unemployment beyond the recessions of the early 1980s and early 1990s. This persistence in unemployment is known as 'hysteresis'.
7. In the late 1990s and early 2000s, equilibrium unemployment fell as labour markets became more flexible and as the lagged effects of the recessions of the early 1980s and early 1990s faded.
8. After the initial shock of the global financial crisis, unemployment fell rapidly in the UK in the 2010s, in part due to labour market flexibility. But underemployment rose and rates of youth and long-term unemployment remained high and were subsequently exacerbated by the pandemic.

12.2 INFLATION

Price levels and inflation rates

Inflation refers to rising price levels; deflation refers to falling price levels. The annual rate of inflation measures the annual percentage *increase* in prices. If the rate of inflation is negative, then prices are falling and we are effectively measuring the rate of deflation.

Typically inflation relates to *consumer* prices. Countries publish consumer price indices each month (see Box 12.3 for more on the UK measures), and the annual rate of inflation is the percentage increase in these indices over the previous 12 months.

A broader measure of inflation relates to the rate at which the prices of all domestically produced goods and services are changing. The price index used in this case is the *GDP deflator* (see page 274 and Case Study 9.16 on the student website). Because it is the price measure of domestic product, the GDP deflator is the price level we plot on the vertical axis of the *AD*/*AS* model (see Section 10.1).

Figure 12.5 shows inflation rates using the GDP deflator for a selection of countries. As you can see, inflation was particularly severe in the mid-1970s, but rates have been relatively low in more recent years with Japan experiencing prolonged periods of falling prices.

You will also find rates of inflation reported for a variety of goods and services. For example, indices are published for commodity prices (see Box 2.3), for food prices, for house prices (see Box 2.1), for import prices, for prices after taking taxes into account and so on. Their respective rates of inflation are simply their annual percentage increase. Likewise, it is possible to give the rate of inflation of wage rates ('wage inflation').

Figure 12.6 shows four inflation rate measures for the UK from 2001. Of these the annual growth of average weekly earnings shows the most variability. This has implications for the purchasing power of workers. We can see, for example, a marked erosion of purchasing power in 2009 following the financial crisis and again, to some degree, during 2014.

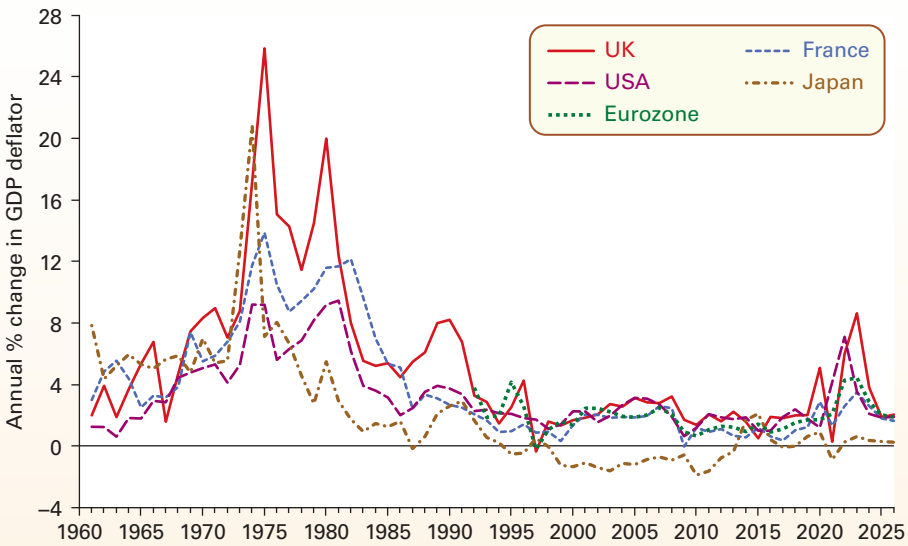
TC 5 p 17

Before we proceed, a word of caution: be careful not to confuse a rise or fall in the rate of *inflation* with a rise or fall in *prices*. An increase in the rate of inflation means a *faster* increase in prices. A fall in the rate of inflation means a *slower* increase in prices (but still an increase as long as the rate of inflation is positive).

Introduction to the causes of inflation

We can identify two principal types of inflation: demand-pull and cost-push inflation. Demand-pull inflation is caused by continuing rises in aggregate demand while cost-push inflation is caused by continuing rises in costs, which occur *independently* of aggregate demand. We can use our aggregate demand and supply framework to analyse demand-pull and cost-push inflation.

Figure 12.5 Economy-wide inflation rates



Notes: Data from 2022 based on forecasts; inflation rate is the annual rate of increase in the GDP deflator; Eurozone = 19 countries using the euro (as of April 2022)

Source: Based on data in *AMECO Database* (European Commission) (to 1981) and *World Economic Outlook* (IMF, October 2022)

BOX 12.3 **INFLATION AND LIVING STANDARDS** CASE STUDIES & APPLICATIONS

The return of inflation

The rapid rise in prices in the early 2020s saw the issue of inflation make news headlines – the first time for several years. But why is inflation a problem? What effect does it have on living standards?

Purchasing power

Purchasing power is fundamental to people's living standards. This can be understood by revisiting the distinction between nominal and real values. When the *nominal* (money) incomes of households grow less quickly than consumer prices, households experience a decline in their *real* incomes: the purchasing power of their incomes is eroded.

TC 5 p 17

The chart shows the rate of CPIH inflation alongside the annual growth of average weekly earnings. It shows an erosion of purchasing power from 2009 following the financial crisis, which was to persist through much of the first half of the 2010s, and again, to some degree, during 2017 and 2020. Then, in 2021–22 households experienced a marked erosion of real incomes. The period became synonymous with a cost-of-living crisis as the prices of energy, food and other items soared.

My inflation is not your inflation

Official measures of inflation capture changes in the cost of a *representative* shopping basket. However, the experiences and effects of inflation inevitably vary across people and business.

Consider now how sharp rises in energy costs in the early 2020s led to a different experience of inflation across different groups of households and firms.

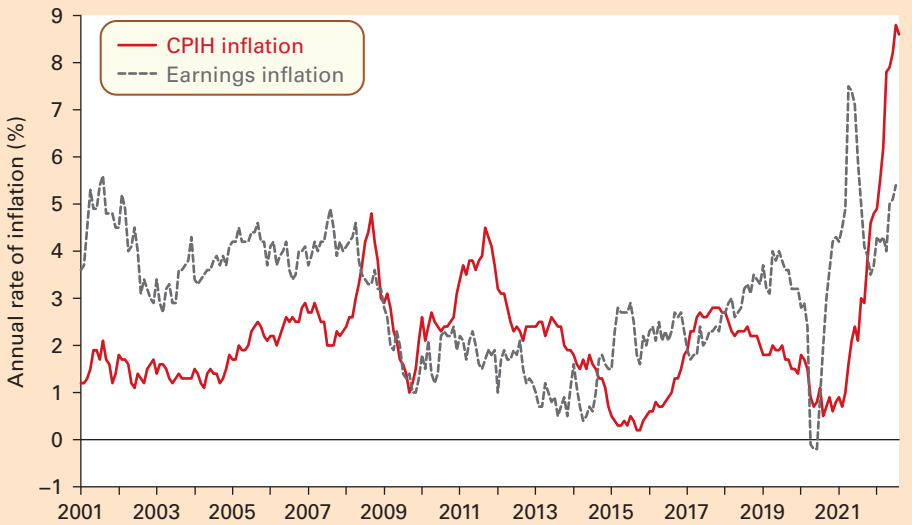
Households. Rising energy costs disproportionately affect low-income households because such costs make up a larger share of their budgets. The effect was to increase the number of households facing fuel poverty. This occurs when poorer households are left with disposable income below the poverty line (an income less than 60 per cent of the national median) once housing costs and energy needs have been met. 'Cutting back on eating or heating' became the agonising choice of many poor people.

Firms. In Box 5.4 (see pages 100–1) we introduced the concept of the economic vulnerability of firms. Type 2 vulnerability relates to a firm's reliance on external or bought-in inputs. Firms that are heavy users of energy, including those in energy-intensive manufacturing and transport, saw large upwards shifts of their average cost (*AC*) curves, with significant implications for their profitability.

Do you personally lose more or less than the average person from inflation? Why?

Draw up a list of those who are most likely to gain and those who are most likely to lose from inflation.

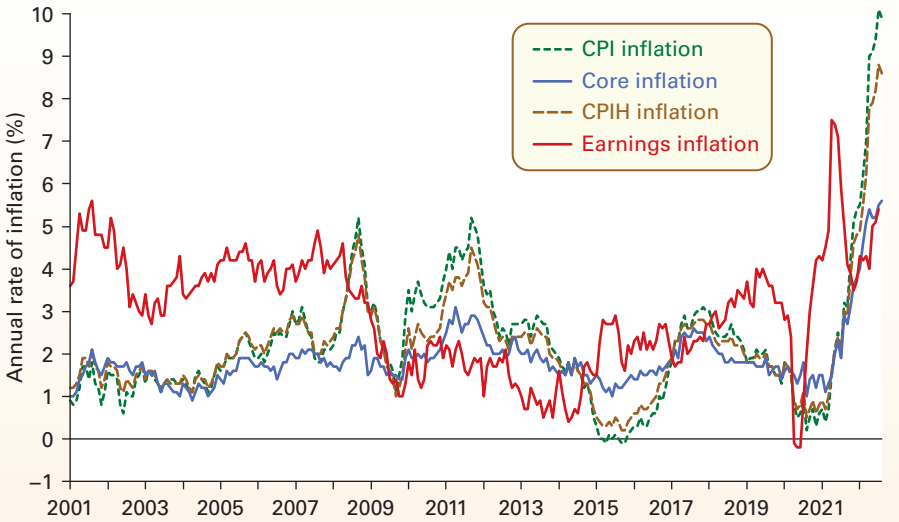
Consumer price inflation and earnings growth



Notes: CPIH is CPI plus owner-occupied housing costs and council tax; Earnings inflation is the annual growth in monthly-average weekly regular pay (excluding bonuses and arrears)

Source: Based on *Time Series Data,* series L550 and K8IA (ONS)

Figure 12.6 Selection of annual UK inflation rates



Notes: CPIH is CPI plus owner-occupied housing costs and council tax; Core inflation is CPIH inflation excluding energy, food, alcoholic beverages and tobacco; Earnings inflation is the annual growth in monthly-average weekly regular pay (excluding bonuses and arrears)

Source: Based on *Time Series Data*, series D7G7, L55O, L5LQ, and K8IA (ONS)

Demand-pull inflation

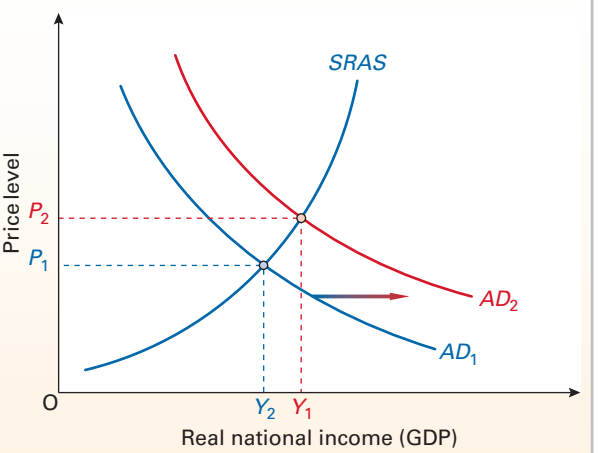
Demand-pull inflation can be represented by *continuous* shifts of the *AD* curve to the right. In Figure 12.7, the *AD* curve shifts to the right and continues doing so. Firms will respond to the rise in aggregate demand partly by raising prices and partly by increasing output (there is a move up along the *SRAS* curve). Just how much they raise prices depends on how much their costs rise as a result of increasing output. This in turn depends on how close actual output is to potential output. The less slack there is in the economy, the more firms will respond to a rise in demand by raising their prices: in other words, the steeper the *SRAS* curve, the larger the rise in prices.

Sometimes there may be a *single* increase in demand (or a 'demand shock') such as new government expenditure on specific projects. The effect is to give a *single* rise in the price level. Although this causes inflation in the short run, once the effect has taken place, inflation will fall back to zero. For inflation to persist, there must be *continuing* increases in aggregate demand and thus continuing rises in the price level. If inflation is to rise, the rate of increase in aggregate demand must also rise.

Demand-pull inflation is typically associated with a booming economy. Many economists therefore argue that it is the counterpart of demand-deficient (cyclical) unemployment. When the economy is in recession, demand-deficient unemployment will be high, but demand-pull inflation will be low. When, on the other hand, the economy is near the peak of the business cycle, demand-pull inflation will be high, but demand-deficient unemployment will be low.

TC 3 p 10

Figure 12.7 Demand-pull inflation



Definition

Demand-pull inflation Inflation caused by persistent rises in aggregate demand.

Cost-push inflation

Cost-push inflation is associated with continuing rises in costs and hence continuing leftward (upward) shifts in the *SRAS* curve. Such shifts occur when costs of production rise *independently* of aggregate demand. If firms face a rise in costs, they will respond partly by raising prices and passing the costs on to the consumer, and partly by cutting back on production. This is illustrated in Figure 12.8. There is an upward shift in the aggregate supply curve: from $SRAS_1$ to $SRAS_2$. This causes the price level to rise to P_2 and the level of output (real GDP) to *fall* to Y_2.

Pause for thought

If there is a general rise in costs of production across the country, does this necessarily mean that there is pure cost-push inflation?

Just how much firms raise prices and cut back on production depends on the shape of the aggregate demand curve. The less elastic the *AD* curve, the less sales will fall as a result of any price rise, and hence the more firms will be able to pass on the rise in their costs to consumers as higher prices.

Note that the effect on output and employment is the opposite of demand-pull inflation. With demand-pull inflation, output and hence employment tend to rise because of the underlying rise in demand. With cost-push inflation, however, output and employment tend to fall.

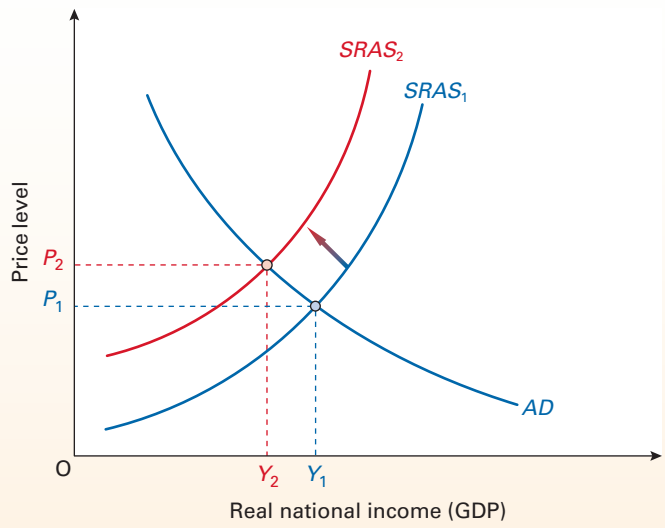
As with demand-pull inflation, we must distinguish between *one-off* increases in cost (a 'supply shock') and *continuing* increases. If there is a one-off increase in costs, there will be a one-off rise in the price level. For example, if the government raises the excise duty on petrol and diesel, there will be a single rise in fuel prices and hence in firms' fuel costs. This will cause *temporary* inflation while the price rise is passed on through the economy. Once this has occurred, prices will stabilise at the new level and the rate of inflation will fall back to the level it was again. If cost-push inflation is to continue over a number of years, therefore, costs must *continually* increase. If cost-push inflation is to *rise*, the rate of increase in costs must also rise.

Rises in costs may originate from a number of different sources, such as trade unions pushing up wages, firms with monopoly power raising prices in order to increase their profits, or increases in international commodity prices. With increased international competition and the stability of inflationary expectations, cost-push pressures generally decreased. Instead, small variations in inflation rates came to reflect the volatility of global commodity prices. However, this changed from 2021 when a combination of supply-side factors, including rising shipping costs and commodity prices, saw inflation rates rise globally. One concern was that the longer the inflationary shock persisted, the more likely that inflationary expectations would rise, so contributing to the persistence of high inflation rates.

Definition

Cost-push inflation Inflation caused by persistent rises in costs of production (independently of demand).

Figure 12.8 Cost-push inflation



BOX 12.4 **COST-PUSH INFLATION** EXPLORING ECONOMICS

Cost-push inflation and supply shocks

It is important to distinguish a *single* supply shock, such as a rise in oil prices or an increase in VAT, from a *continuing* upward pressure on costs, such as workers continually demanding increases in real wages above the level of labour productivity, or firms continually using their monopoly power to increase the real value of profits.

A single supply shock will give a single upward movement in the *SRAS* curve. Prices will move to a new higher equilibrium. An example occurred in the UK in January 2011 with the rise in VAT from 17.5 to 20 per cent. Cost-push inflation in this case is a temporary phenomenon. Once the new higher price level has been reached, the cost-push inflation disappears.

Periods of *continuous* upward pressure on costs result in repeated shifts of the *SRAS* curve. The early 2020s were characterised by a series of cost-inflationary pressures. These included the supply-chain problems that faced some sectors as economies opened up at different speeds from the COVID-19 pandemic, the associated rise in shipping costs (which is important because over 80 per cent of traded goods are transported by sea) and the increase in commodity prices that were exacerbated by the Russian invasion of Ukraine.

A concern for policymakers was that the persistence of inflation arising from the inflationary shocks would be worsened by inflationary expectations. Workers would demand higher wages to compensate for the rise in the cost of living; firms would put up prices to cover the higher costs, expecting that competitors would do the same.

Inflationary shocks, such as those in the 2020s, can create a policy dilemma for central banks that are charged with meeting inflation-rate targets. Should they raise interest rates to slow the growth in aggregate demand, even if much of the cause of consumer price inflation is on the supply side? Although this will help to reduce inflationary expectations, it can lead to a recession.

If consumer demand rises and firms respond by raising prices, is this necessarily an example of demand-pull inflation? Could it be cost-push inflation – at least in part? (Clue: why might consumer demand have risen?)

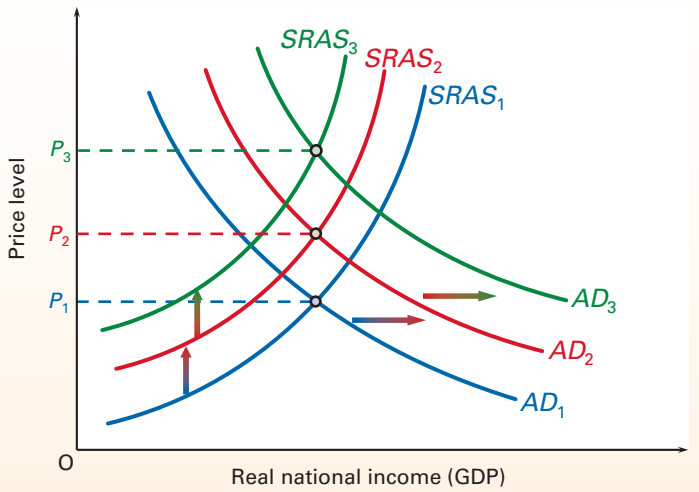
Conduct a search of news items over recent months to identify possible examples of 'genuine' cost-push inflation.

The interaction of demand-pull and cost-push inflation

Demand-pull and cost-push inflation can occur together, since wage and price rises can be caused both by increases in aggregate demand and by independent causes pushing up costs. Even when an inflationary process *starts* as either demand-pull or cost-push, it is often difficult to separate the two. An initial cost-push inflation may encourage the government to expand aggregate demand to offset rises in unemployment. Alternatively, an initial demand-pull inflation may strengthen the power of certain groups, who then use this power to drive up costs. Either way, the result is likely to be continuing rightward shifts in the *AD* curve and upward shifts in the *SRAS* curve. Prices will carry on rising. This is illustrated in Figure 12.9.

TC 3 p 10

Figure 12.9 The interaction of demand-pull and cost-push inflation



Expectations and inflation

TC 11 p 61

Workers and firms take account of the *expected* rate of inflation when making decisions.

Imagine that a union and an employer are negotiating a wage increase. Let us assume that both sides expect a rate of inflation of 2 per cent. The union will be happy to receive a wage rise somewhat above 2 per cent. That way, the members would be getting a *real* rise in incomes. The employers will be happy to pay a wage rise somewhat below 2 per cent. After all, they can put their price up by approximately 2 per cent, knowing that their rivals will do the same. The actual wage rise that the two sides agree on will thus be somewhere around 2 per cent.

Now let us assume that the expected rate of inflation is 5 per cent. Both sides will now negotiate around this benchmark, with the outcome being somewhere around 5 per cent.

Thus the higher the expected rate of inflation, the higher the level of pay settlements and price rises will be, and hence the higher the resulting actual rate of inflation will be.

The importance of expectations of inflation is examined in Sections 12.4 and 12.5.

Recap

1. Demand-pull inflation occurs as a result of continuing increases in aggregate demand.
2. Cost-push inflation occurs when there are continuing increases in the costs of production independent of rises in aggregate demand. Single 'supply shocks' will lead to one-off increases in prices but not to continuing inflation.
3. Expectations play a crucial role in determining the rate of inflation. The higher people expect inflation to be, the higher it will be.

12.3 THE RELATIONSHIP BETWEEN OUTPUT, UNEMPLOYMENT AND INFLATION: THE SHORT RUN

What is the relationship between output, unemployment and inflation?

Fluctuations in aggregate demand and aggregate supply generate business cycles, with fluctuations in the level of output (real national income) and the rates of unemployment and inflation.

Our macroeconomic models allow us to analyse how these key macroeconomic variables fluctuate, and the relationships between them. These models are, in effect, different windows looking out on to the economy.

We begin with the *AD/AS* model before considering the Phillips curve.

The *AD/AS* model

The flexibility of aggregate supply

As we saw in Chapter 10, there are different perspectives on the flexibility of aggregate supply in response to demand fluctuations. This is important because the balance between output and price changes in response to shifts in the *AD* curve and depends on the slope of the *SRAS* curve.

The extreme Keynesian position. Consider again the extreme Keynesian short-run aggregate supply curve that underpins the simple Keynesian model (see Section 9.5). The curve is *horizontal* up to the full employment output level, Y_F and is labelled as $SRAS_1$ in Figure 12.10. Up to Y_F, output and employment can rise with no rise in prices at all. As this happens, the deflationary gap is closed. At Y_F no further rises in output are possible.[1]

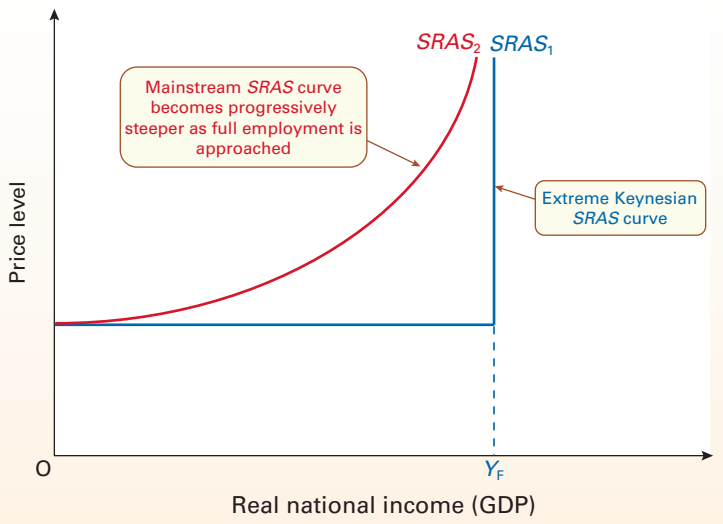
Any further rise in aggregate demand is entirely reflected in higher prices. Instead, an inflationary gap opens. This implies that either inflation *or* unemployment can occur, but not both simultaneously.

The mainstream view. Now consider the mainstream or orthodox *SRAS* curve. This is represented by $SRAS_2$ in Figure 12.10. It gets progressively steeper, reflecting the likelihood that, as national output increases, marginal costs rise more and more rapidly as capacity constraints and bottlenecks increasingly affect firms. A rise in aggregate demand can therefore lead to *both* a reduction in unemployment *and* a rise in prices at output levels *below* Y_F.

Because inflation can occur *before* the full-employment level of real national income is reached, increases in (nominal) aggregate demand no longer increase real national income by the full extent of the multiplier process. Since the progressively steeper

[1] For simplicity, we are ignoring here any temporary rise in output above normal capacity.

Figure 12.10 Alternative *SRAS* curves



SRAS curve means that inflationary pressures become more significant as national output rises, the size of the multiplier grows smaller, and the closer the economy is to the full-employment national income level. At higher levels of national output, an increasingly large part of any increase in nominal aggregate demand is reflected in higher prices and an increasingly small part in higher output.

TC 5 p 17

The flexibility of aggregate supply is therefore crucial in determining the impact of fluctuations in aggregate demand on unemployment and inflation. However, it is important to recognise other influences on unemployment and inflation.

Other types of inflation and unemployment

First, there are *other* types of inflation not caused by an excess of aggregate demand, such as cost-push and expectations-generated inflation. Box 12.5 looks at the actual relationship between the rate of inflation and the amount of excess or deficient demand, as measured by output gaps, in the UK since 1985.

Second, there are *other* types of unemployment not caused by a lack of aggregate demand. Examples include frictional and structural unemployment.

Thus, even if a government could manipulate national income so as to achieve a zero-output gap, this would not eliminate all inflation and unemployment – only demand-pull inflation and demand-deficient unemployment.

Keynesians argue, therefore, that governments should use a whole package of policies, each tailored to the specific type of problem. But certainly, one of the most important of these policies will be the management of aggregate demand.

The goods and labour markets

The short-run relationship between output, unemployment and prices is closely tied to the labour market. We can illustrate this by combining the labour market diagram (Figure 12.5) with our *AD*/*AS* diagram, as in Figure 12.11. (We adopt the mainstream upward-sloping *SRAS* curve.)

Assume that the economy's equilibrium output level (real national income) is Y_1, as shown in Figure 12.11(b), where aggregate demand and supply (for goods) are equal. Assume too that equilibrium in the goods market corresponds with equilibrium in the labour market.

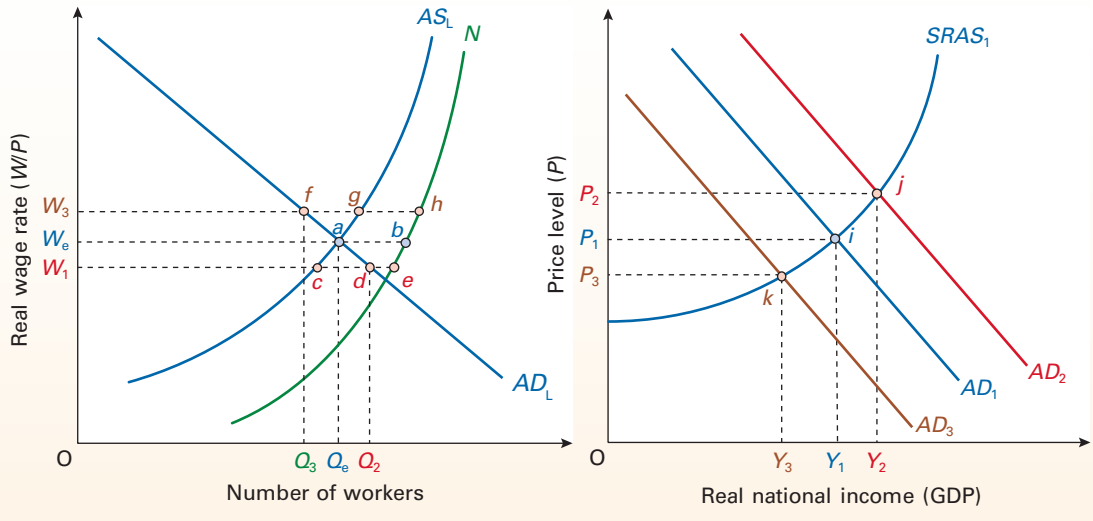
In Figure 12.11(a) the equilibrium real wage is shown as W_e, where $AD_L = AS_L$, with Q_e workers employed. There is no disequilibrium unemployment at this real wage rate, but there is some equilibrium or natural unemployment. As we saw in Section 12.1, this is largely made up of frictional and structural unemployment. It is represented by the distance $(b - a)$. Thus, even when the goods and labour markets are equilibrium there is unemployment.

TC 3 p 10

A rise in aggregate demand. Now assume that aggregate demand for goods rises. This is shown by a rightward move of the aggregate demand curve from AD_1 to AD_2. It causes a movement up the short-run aggregate supply curve, $SRAS_1$, from point *i* to point *j*. The price level rises from P_1 to P_2.

KI 8 p 24

Figure 12.11 The labour market and output determination



To see how the labour market has facilitated this increase in output, consider now the labour market diagram (Figure 12.11(a)). The rise in the price level, causes the real wage to fall below W_e, say to W_1. At real wage rate W_1, labour demand is Q_2. If we assume that, at least in the short run, workers supply this additional labour demand, perhaps because of contractual obligations, then employment has risen from Q_e to Q_2. Unemployment is now represented by the distance $(e - d)$.

The issue that should now be apparent from looking at Figure 12.11(a) is that there is an excess demand for labour of $d - c$. This would be expected to drive up the money (nominal) wage rate until the *real* wage rate has returned to W_e, and employment to its equilibrium level of Q_e. As this occurs, the *SRAS* curve will shift upwards to the left, offsetting in the process the increase in *AD* until output is again at Y_1.

TC 5 p 17

An important determinant of the short-run relationship between economic activity and inflation in the face of fluctuating aggregate demand is therefore how quickly wages adjust. How flexible are wages?

KI 8 p 24

A fall in aggregate demand. The flexibility of wages, some argue, is even more significant when aggregate demand falls. In this case, the *AD* curve in Figure 12.11(a) shifts leftwards from AD_1 to AD_3, output falls from Y_1 to Y_3 and the price level from P_1 to P_3.

The fall in the price level causes the real wage to rise above W_e to W_3 in Figure 12.11(a). We now have an excess supply of labour. Employment is constrained by the fall in demand in the goods markets. Employment thus falls to Q_3 and unemployment is represented by the distance $(h - f)$.

Many Keynesians argue that an economy could get stuck at a low level of output, such as Y_3 in Figure 12.11(b) following a fall in aggregate demand, if, as might be expected, there is resistance from workers to accept cuts in money wage rates. Governments, they argue, can play an important role in helping to stimulate aggregate demand.

Pause for thought

What are the levels of demand-deficient unemployment and equilibrium unemployment at a real wage rate of W_3 in Figure 12.11(a)?

The Phillips curve

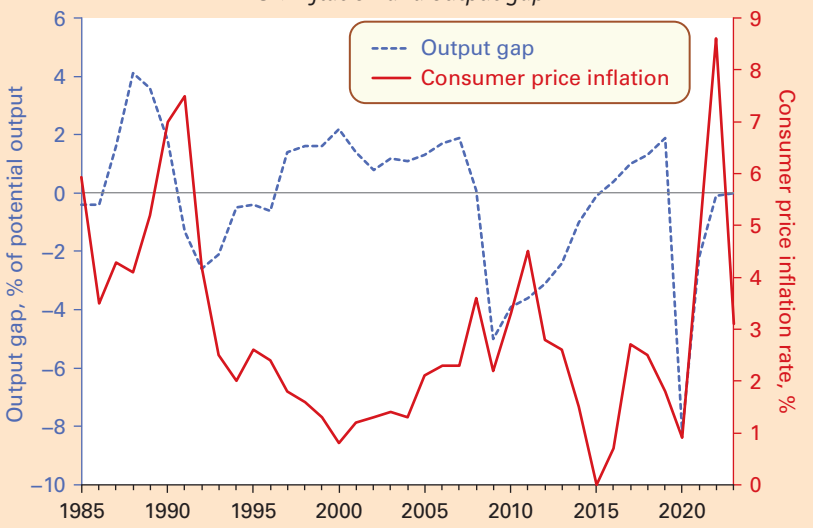
The relationship between inflation and unemployment was examined in a famous article by A.W. Phillips back in 1958 (see Case Study 12.9 on the student website). He showed the statistical relationship between wage inflation and unemployment in the UK from 1861 to 1957. With wage inflation on the vertical axis and the unemployment rate on the horizontal axis, a scatter of points was obtained. Each point represented the observation for a particular year. The curve that best fitted the scatter has become known as the **Phillips curve**. It is illustrated in Figure 12.12 and shows an inverse relationship between inflation and unemployment.

BOX 12.5 MIND THE GAP

CASE STUDIES AND APPLICATIONS

Do output gaps explain inflation?

UK inflation and output gap



Notes: Data from 2022 based on forecasts

Source: Based on data in *AMECO Database* (European Commission, DGECFIN, May 2022) and various forecasts

As we saw in Chapter 9 (see page 240 and Box 9.10), the output gap measures the difference between an economy's actual level of output and its potential or normal-capacity output. A positive output gap shows that the level of actual output is *greater* than the potential level, while a negative output gap shows that the level of output is below the potential level.

The magnitude of the output gap, which is usually expressed as a percentage of potential output, enables us to assess the extent of any demand deficiency (negative output gap) or excess demand (positive output gap).

The 'mainstream view' of the slope of the short-run aggregate supply (see $SRAS_2$ in Figure 12.10) is that it is determined by the amount of slack in the economy. As the economy approaches or exceeds its potential output the aggregate supply curve becomes steeper as firms' marginal costs rise faster. Consequently, increases in demand at output levels close to or in excess of an economy's potential output will exert more upward pressure on prices than if the economy has a more significant amount of slack. This suggests that the rate of price inflation is positively related to the size of the output gap.

The chart plots the output gap (as a percentage of potential output) and the annual rate of consumer price inflation for the UK since 1985.

In the late 1980s there is evidence of a positive, albeit lagged, correlation between output gaps and inflation rates. In other words, changes in the rates of inflation lagged those in the size of output gaps. This is because it took time for price pressures to work fully through the economy. This reflects, in part, the fact that some prices, such as wage rates, are adjusted periodically.

Since the 1990s, however, the relationship between output gaps and inflation rates is less clear. Indeed, much of the period up to the early 2020s was characterised by relative low rates of inflation regardless of the size of output gaps. Then, as COVID restrictions started to be eased, the global economy saw the emergence of significant inflationary pressures emanating not just from the rise in demand, but from the supply-side that were then exacerbated by the Russian invasion of Ukraine. This demonstrates that, in addition to the output gap, there are other influences on inflation rates.

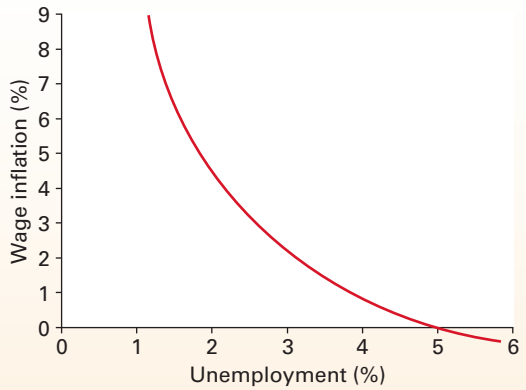
Cost-push pressures. There were significant exogenous cost shocks. Large rises in commodity prices, such as crude oil, natural gas, metals, minerals and staple crops, were only partly caused by the rise in demand. They were largely due to supply problems and the Ukrainian war. The IMF price index for energy-related commodities rose by 100 per cent in the year to March 2022, while that for non-energy commodities rose by over 30 per cent.

Inflation rate expectations. The adoption of clear and credible inflation targets by central banks helps to anchor price and wage setting to the inflation rate target, irrespective of the state of the economy. In 2022, however, despite soaring inflation, central banks were somewhat slow to raise interest rates for fear of causing a recession. This caused a rise in the expected rate of inflation. People were no longer convinced that central banks would act to keep inflation in check – at least in the short term.

1. *What factors may have resulted in the generally lower inflation rates experienced by the UK from the 1990s?*
2. *Do credible inflation rate targets guarantee low inflation?*

Download the latest available version of Forecasts for the UK economy, which is a monthly comparison of independent forecasts complied by HM Treasury. Then, compare the average forecast for the annual CPI inflation rate for the periods available with the Bank of England's target of 2 per cent (plus or minus 1 percentage point).

Figure 12.12 The Phillips curve



Given that wage increases over the period were approximately 2 per cent above price increases (made possible because of increases in labour productivity), a similar-shaped, but lower curve could be plotted showing the relationship between *price* inflation and unemployment.

The curve has often been used to illustrate the short-run effects of changes in (real) aggregate demand. When aggregate demand rose (relative to potential output), inflation rose and unemployment fell: there was a movement upwards along the curve. When aggregate demand fell, there was a movement downwards along the curve.

The Phillips curve was bowed in to the origin. The usual explanation for this is that as aggregate demand expanded, at first there would be plenty of surplus labour, which could be employed to meet the extra demand without the need to raise wage rates very much. But as labour became increasingly scarce, firms would find that they had to offer increasingly higher wage rates to obtain the labour they required, and the position of trade unions would be increasingly strengthened.

The *position* of the Phillips curve depended on *non*-demand factors causing inflation and unemployment: frictional and structural unemployment; and cost-push and expectations-generated inflation. If any of these non-demand factors changed so as to raise inflation or unemployment, the curve would shift outwards to the right.

Pause for thought

What would an alternative Phillips curve drawn showing the relationship between inflation and output *look like?*

Definition

Phillips curve A curve showing the relationship between (price) inflation and unemployment. The original Phillips curve plotted *wage* inflation against unemployment for the years 1861–1957.

The Phillips curve seemed to present governments with a simple policy choice. They could trade off inflation against unemployment. Lower unemployment could be bought at the cost of higher inflation, and vice versa. Unfortunately, the experience since the late 1960s has suggested that no such simple relationship exists beyond the short run.

KI 31 p 206

The breakdown of the Phillips curve

From about 1967 the Phillips curve relationship seemed to break down. The UK, along with many other countries in the Western world, began to experience growing unemployment and higher rates of inflation as well.

Figure 12.13 shows consumer price inflation and unemployment in the UK from 1960. From 1960 to 1967 a curve similar to the Phillips curve can be fitted through the data. From 1968 to the early 1990s, however, no simple picture emerges. Certainly the original Phillips curve could no longer fit the data; but whether the curve shifted to the right and then back again somewhat (the broken lines), or whether the relationship broke down completely, or whether there was some quite different relationship between inflation and unemployment, is not clear by simply looking at the data.

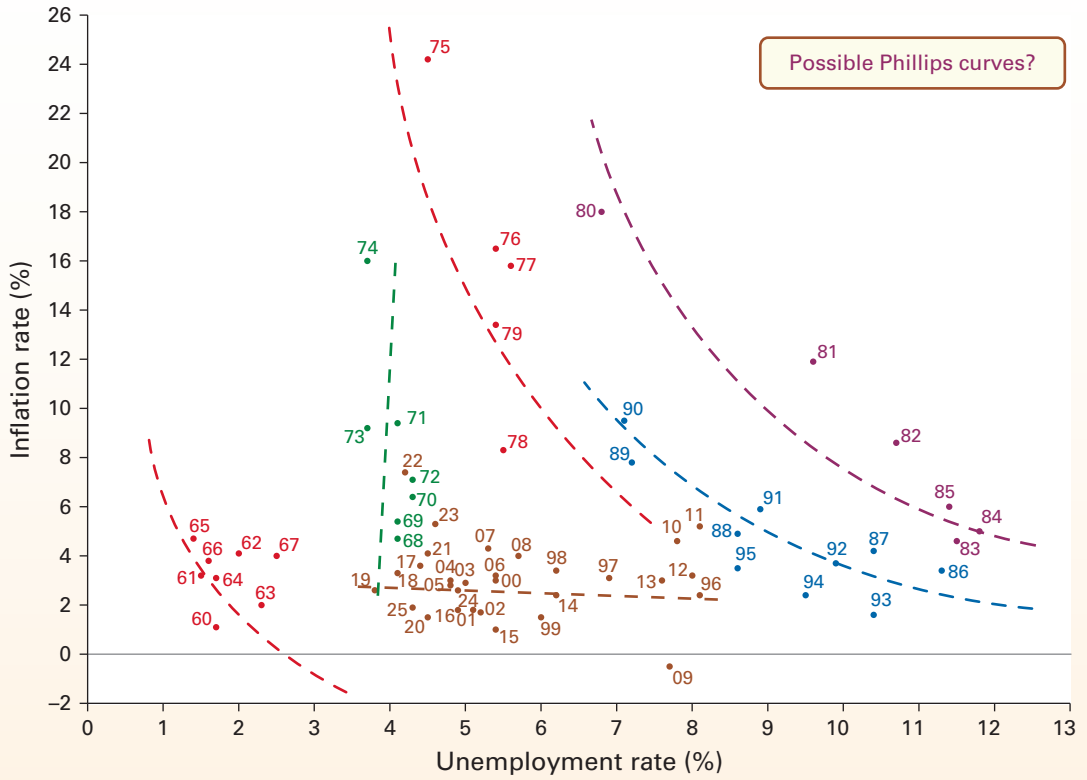
Since 1997 the Bank of England has been targeting consumer price inflation (see Section 12.5). For much of this period, the 'curve' would seem to have become a virtually horizontal straight line!

However, from the late 2000s, against a backdrop of significant economic volatility, including two global economic shocks, the range of inflation rates increased. This was driven by supply-side factors, including the volatility of commodity prices (see Box 2.3). The extent of the inflationary shocks from 2021/22 led some to liken the effects on the macroeconomic environment to those in the 1970s.

Nonetheless, despite difficulties in keeping inflation to target, many economists continue to argue that expectations concerning future inflation rates are a crucial influence on the inflation-unemployment relationship. In the next section, we look at how expectations have been incorporated into the Phillips curve relationship.

TC 11 p 61

Figure 12.13 The breakdown of the Phillips curve



Source: Based on data from the Office for National Statistics; forecasts based on data from *World Economic Outlook Database* (IMF)

Recap

1. The *SRAS* curve in the simple Keynesian model is horizontal up to Y_F and then vertical. Therefore, up to this point output can expand or contract, and unemployment rise or fall, without affecting the economy's price level.
2. The mainstream *SRAS* is upward sloping but getting steeper as full employment is approached and as bottlenecks increasingly occur. Therefore, inflation can occur *before* Y_F or potential output (the economy's normal-capacity output level).
3. With an upward sloping *SRAS* curve, increasing levels of aggregate demand will no longer increase real national income by the full extent of the multiplier process: increasingly large parts of increases in aggregate demand are reflected in higher prices and a smaller part in higher output.
4. The labour market is an important determinant of the relationship between output, unemployment and inflation. In the short term, with contractual obligations and sticky wages, changes in aggregate demand are likely to affect the quantity of workers employed and hence national output.
5. The less quickly that money wages adjust to restore the equilibrium real wage rate, the larger and more enduring output changes are likely to be. Some economists argue that money wages may be less flexible in response to *falling* aggregate demand with workers resistant to accepting lower money wages. Therefore, government action may need to boost aggregate demand.
6. The Phillips curve showed the trade-off between inflation and unemployment. There seemed to be a simple inverse relationship between the two. After 1967, however, the relationship broke down as inflation *and* unemployment rose.
7. Today, many central banks target inflation. The effect has been to make the Phillips curve appear more like an almost horizontal straight line. However, the instability of the macroeconomic environment since 2007, driven by two global economic shocks, has seen a greater variability of inflation rates, with the potential to destabilise inflationary expectations.

12.4 THE RELATIONSHIP BETWEEN INFLATION AND UNEMPLOYMENT: INTRODUCING EXPECTATIONS

What happens when people come to expect inflation?

A major contribution to the theory of unemployment and inflation was made by Milton Friedman and others in the late 1960s. They incorporated people's expectations about the future level of prices into the Phillips curve.

TC 11 p 61

Friedman argued that the theoretical trade-off implied by the simple Phillips curve relationship relied on permanent changes in the *real* wage. For the rate of unemployment to remain below its equilibrium rate, workers would need to consistently supply labour at *below* the equilibrium real wage rate (e.g. W_e in Figure 12.11 on page 359). He argued that this was implausible other than in the short term and that wage inflation would surely catch up with consumer price inflation to restore the real wage equilibrium.

TC 5 p 17

Friedman argued that the Phillips curve relationship needed to be supplemented or augmented with inflationary expectations.

The expectations-augmented Phillips curve

In its simplest form, the **expectations-augmented Phillips curve** (**EAPC**) may be expressed as:

$$\pi = f\left(\frac{1}{U}\right) + \pi^e + k$$

This states that the rate of price inflation (π) depends on three things.

- First, it is a function (f) of the inverse of unemployment ($1/U$). This is simply the original Phillips curve relationship. A rise in aggregate demand will lead to a fall in unemployment (a rise in $1/U$) and a rise in inflation – a move *along* the EAPC.
- Second, the expected rate of inflation π^e must be added to the inflation that would result simply from the level of excess demand represented by ($1/U$).

TC 11 p 61

- Third, if there are any exogenous cost pressures on inflation (k) (such as increases in international commodity prices), this must be added too.

Thus if people expected a 3 per cent inflation rate ($\pi^e = 3\%$) and if excess demand were causing demand-pull inflation of 2 per cent ($f(1/U) = 2\%$) and exogenous increases in costs were adding another 1 per cent to inflation ($k = 1\%$), actual inflation would be $3 + 2 + 1 = 6$ per cent.

By augmenting the Phillips curve with expectations, we move from a single curve to a family of Phillips curves. The vertical position of each curve is determined by inflationary expectations (π^e) and any exogenous costs pressures (k). If we assume that on average $k = 0$, then we can simplify the *EAPC* framework so that each expectations-augmented Phillips curve is associated with a particular rate of inflationary expectations.

Figure 12.14 illustrates two hypothetical expectations-augmented Phillips curves. $EAPC_1$ is drawn for an expected inflation rate π_1, while $EAPC_2$ is drawn for an expected inflation rate π_2. Hence, for any given rate of unemployment, $EAPC_2$ is vertically higher than $EAPC_1$ by the difference $(\pi_2 - \pi_1)$. At unemployment rate U_1, for example, the difference between the corresponding inflation rates for $EAPC_2$ and $EAPC_1$ of $(\pi_b - \pi_a)$ is equivalent to $(\pi_2 - \pi_1)$.

Natural rate hypothesis

A key conclusion of monetarist thinking was that in the long run the equilibrium rate of unemployment is determined independently of aggregate demand. Unemployment will return to its equilibrium rate (and output to its potential level) following fluctuations in aggregate demand. This is known as the **natural rate hypothesis** because Friedman and other monetarists refer to the equilibrium rate as the **natural rate** (U_n). The natural rate of unemployment is sometimes also known as the **non-accelerating-inflation rate of unemployment** (**NAIRU**).

Under the natural rate hypothesis. the long-run Phillips curve (*LRPC*) is vertical. In the long run it is supply-side policies that affect levels of structural and frictional unemployment. However, not all economists accept that the natural rate of unemployment is unaffected by fluctuations in aggregate demand. Keynesian economists, for example, point to the

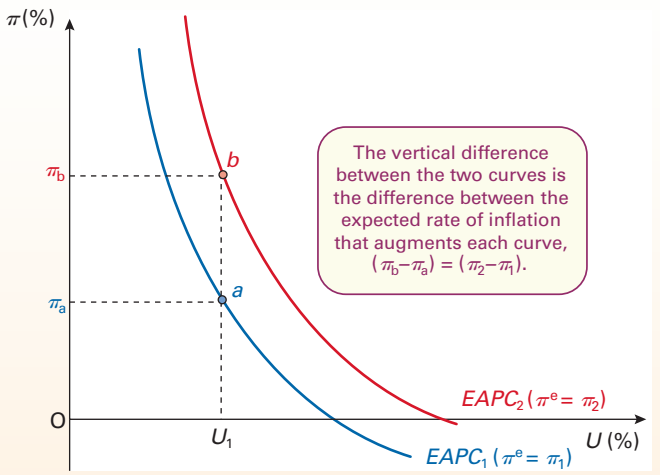
Definitions

Expectations-augmented Phillips curve A (short-run) Phillips curve whose position depends on the expected rate of inflation.

Natural rate hypothesis The theory that, following fluctuations in aggregate demand, unemployment will return to a natural rate. This rate is determined by supply-side factors, such as labour mobility.

Natural rate of unemployment or non-accelerating-inflation rate of unemployment (**NAIRU**) The rate of unemployment consistent with a constant rate of inflation: the rate of unemployment at which the vertical long-run Phillips curve cuts the horizontal axis.

Figure 12.14 Expectations of inflation and the Phillips curve



negative impact that demand deficient unemployment can have on human capital. These and other effects, they argue, can result in a non-vertical *LRPC* (and *LRAS* curve), as we shall see below.

Pause for thought

How might fluctuations in the rate of unemployment affect potential output?

Short-run trade-off

However, in the short run unemployment may deviate from its natural rate. To understand why, we develop a simple monetarist model of the economy sometimes known as the 'fooling model' for reasons that will become clear.

This model is based on two key assumptions. First, prices adjust relatively quickly to ensure equilibrium between demand and supply in both the goods and labour markets. Second, people form **adaptive expectations** of inflation. This occurs when people base their expectations of inflation on *past* inflation rates. If, for example, last year people under-predicted the rate of inflation, then this year they will adapt by revising, i.e. adapting, their expectations of inflation upwards.

Pause for thought

Can it ever be rational to form adaptive expectations?

We will assume that people adopt the simplest form of adaptive expectations: they use last year's actual inflation rate (π_{t-1}) as their prediction for the expected rate of inflation this year (π^e_t). We also assume that the economy's inflation rate last year was zero; no inflation is expected; and there are no exogenous cost pressures on inflation ($k = 0$).

In Figure 12.15 the economy is initially at point *a* with both actual and expected inflation of zero. The goods and labour markets are in equilibrium: $AD = AS$ and unemployment is at its natural rate, U_n.

Increase in aggregate demand. Assume that there is an increase in aggregate demand. The economy moves to, say, point *b* along expectations-augmented Phillips curve, $EAPC_0$. The rate of inflation rises from zero to π_1. But, with people still expecting zero inflation, the labour market experiences a fall in the average real wage rate, leading to a rise in the demand for labour. Assuming that workers are fooled into supplying the additional labour despite the fall in real wages, unemployment falls below its natural rate to U_1.

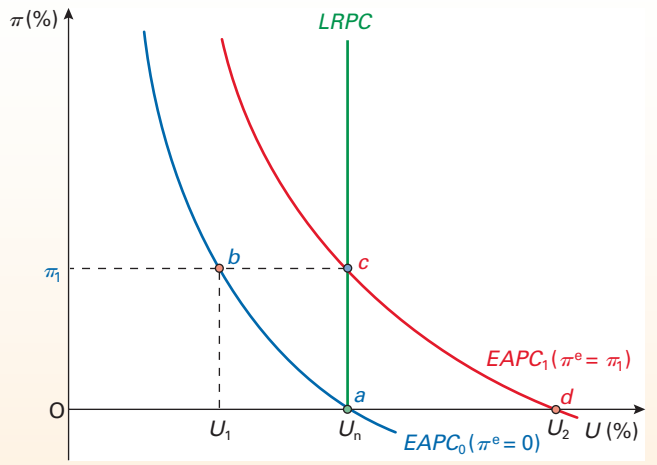
If we now move ahead a year, people will have revised their expectations of inflation upwards to π_1. The result is that the Phillips curve has shifted up vertically by π_1 to $EAPC_1$. If *nominal* aggregate demand (i.e. demand purely in monetary terms, irrespective of the level of prices) continues to rise at the same rate, the whole of the increase will now be absorbed in higher prices. *Real* aggregate demand will fall back to its previous level and the economy will move to point *c* on the long-run Phillips curve. Unemployment will

TC 3 p 10

Definition

Adaptive expectations Where people adjust their expectations of inflation in the light of what has happened to inflation in the past.

Figure 12.15 The fooling model



return to its natural rate, U_n, consistent with the natural rate hypothesis. There is no *demand-pull* inflation now ($f(1/U) = 0$), but inflation is π_1 per cent due to inflationary expectations.

Decrease in aggregate demand. Assume that the economy is at point *c* in Figure 12.15, with expected inflation of π_1. It now experiences a decrease in the growth of nominal demand. Real aggregate demand falls. Let us assume that there is downward pressure on inflation such that the inflation rate falls to zero. The economy moves down $EAPC_1$ to point *d*. In the labour market, the expected average real wage rate increases because the expected inflation rate is π_1. The demand for labour falls and unemployment rises above its natural rate, U_n to U_2.

The following year the expected rate of inflation will fall to zero. The *EAPC* moves vertically down to $EAPC_0$. If the growth in nominal aggregate demand remains at its new lower rate, with inflation now at zero, the economy will again be at point *a* on the long-run Phillips. Unemployment is again at its natural rate; the natural rate hypothesis again holds.

The accelerationist hypothesis

The preceding analysis shows how, when people form adaptive expectations of inflation, unemployment can deviate from its natural rate following changes in aggregate demand, even if markets clear fairly quickly. This raises the theoretical possibility that governments could keep unemployment below the natural rate. But, this would come at a cost, since to do so it must raise nominal aggregate demand at ever increasing rates.

Each time the government is able to raise the nominal growth in aggregate demand there is a transitory period when real aggregate demand rises too. But, inflationary expectations then rise to reflect higher inflation rates. This is mirrored in the labour market by real wages being driven back up to their equilibrium level. Hence, for the government to keep unemployment below its natural rate it needs to keep raising the growth in nominal aggregate demand – in other words, it must increase nominal aggregate demand faster and faster. This, of course, means that nominal aggregate demand needs to grow at more than the rate of inflation. However, the rate of inflation is itself getting progressively higher as people continually raise their inflationary expectations.

The theory that unemployment can be reduced below the natural rate only at the cost of accelerating inflation is known as the **accelerationist hypothesis**. Box 12.6 considers a numerical example of the hypothesis.

New classical perspective

New classical economists go further than the monetarist theory described above. They argue that unless there are unexpected or 'surprise' events, there is no short-run trade-off between inflation and unemployment. The *EAPC* therefore represents a short-run trade-off between inflation and unemployment only in the presence of surprise events.

New classical assumptions

They base their arguments on the two key assumptions we saw in Section 10.2 (see page 281): *continuous*

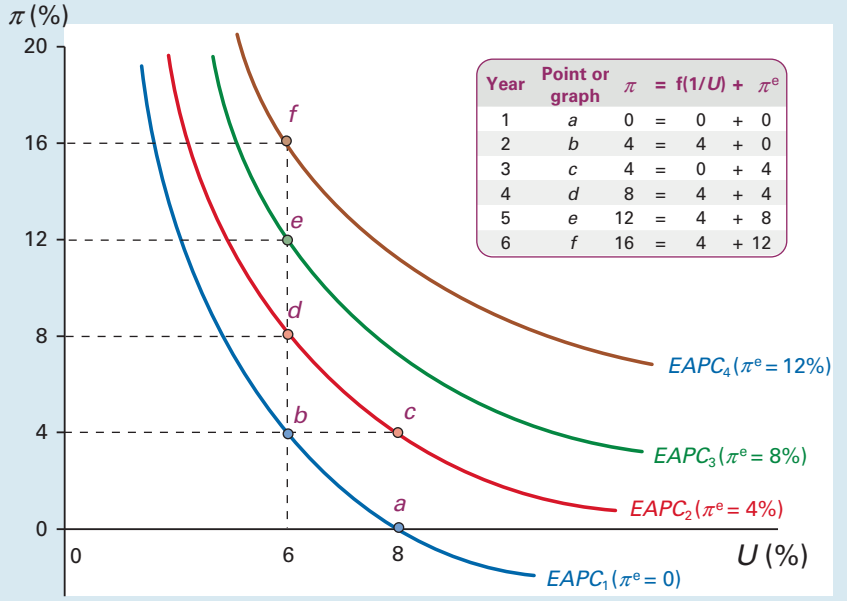
Definition

Accelerationist hypothesis The theory that unemployment can only be reduced below the natural rate at the cost of accelerating inflation.

BOX 12.6 **THE ACCELERATIONIST HYPOTHESIS** EXPLORING ECONOMICS

The race to outpace inflationary expectations

The accelerationist theory of inflation



Year	Point or graph	π	=	f(1/U)	+	π^e
1	*a*	0	=	0	+	0
2	*b*	4	=	4	+	0
3	*c*	4	=	0	+	4
4	*d*	8	=	4	+	4
5	*e*	12	=	4	+	8
6	*f*	16	=	4	+	12

Let us trace the course of inflation and expectations over a number of years in an imaginary economy. To keep the analysis simple, assume there is no growth in the economy and no exogenous cost pressures on inflation ($k = 0$ in the equation on p. 363).

Year 1. Assume that at the outset, in year 1, there is no inflation of any sort; that none is expected; that $AD = AS$; and that equilibrium unemployment is 8 per cent. The economy is at point *a* in the diagram.

Year 2. Now assume that the government expands aggregate demand in order to reduce unemployment. Unemployment falls to 6 per cent. The economy moves to point *b* along $EAPC_1$. Inflation has risen to 4 per cent, but people, basing their expectations of inflation on year 1, still expect zero inflation. There is therefore no shift as yet in the Phillips curve. $EAPC_1$ corresponds to an expected rate of inflation of zero.

Year 3. People now revise their expectations of inflation to the level of year 2. The Phillips curve shifts up by 4 percentage points to $EAPC_2$. If *nominal* aggregate demand (i.e. demand purely in monetary terms, irrespective of the level of prices) continues to rise at the same rate, the whole of the increase will now be absorbed in higher prices. *Real* aggregate demand will fall back to its previous level and the economy will move to point *c*. Unemployment will return to 8 per cent. There is no *demand-pull* inflation now ($f(1/U) = 0$), but inflation is still 4 per cent due to expectations ($\pi^e = 4$ per cent).

Year 4. Assume now that the government expands real aggregate demand again so as to reduce unemployment once more to 6 per cent. This time it must expand nominal aggregate demand by more than it did in year 2, because this time, as well as reducing unemployment, it also has to validate the 4 per cent expected inflation. The economy moves to point *d* along $EAPC_2$. Inflation is now 8 per cent.

Year 5. Expected inflation is now 8 per cent (the level of actual inflation in year 4). The Phillips curve shifts up to $EAPC_3$. If at the same time the government tries to keep unemployment at 6 per cent, it must expand nominal aggregate demand 4 per cent faster in order to validate the 8 per cent expected inflation. The economy moves to point *e* along $EAPC_3$. Inflation is now 12 per cent.

Year 6 onwards. To keep unemployment at 6 per cent, the government must continue to increase nominal aggregate demand by 4 per cent more than the previous year. As the expected inflation rate goes on rising, the Phillips curve will go on shifting up each year.

What determines how rapidly the short-run Phillips curves in the diagram shift upwards?

Construct a table like the one in the diagram, only this time assume that the government wishes to reduce unemployment to 5 per cent. Assume that every year from year 1 onwards the government is prepared to expand aggregate demand by whatever it takes to do this. If this expansion of demand gives $f(1/U) = 7$ per cent, fill in the table for the first six years. Do you think that after a couple of years people might begin to base their expectations differently?

TC 5 p 17

market clearing and *rational expectations*. Because prices and wage rates are flexible, markets clear very rapidly. This means that there will be no disequilibrium unemployment. All unemployment will be equilibrium unemployment, or 'voluntary unemployment' as new classical economists prefer to call it.

In the monetarist model, expectations are adaptive. They are based on *past* information and thus take time to catch up with changes in aggregate demand. Hence, for a short time a rise in nominal aggregate demand will raise output, and reduce unemployment below the equilibrium level, while price and wage increases are still relatively low.

TC 11 p 61

The new classical analysis is based on rational expectations. Rational expectations are not based on past rates of inflation. Instead, they are based on the current state of the economy and the current policies being pursued by the government. Workers and firms look at the information available to them – at the various forecasts that are published, at various economic indicators and the assessments of them by various commentators, at government pronouncements, and so on. From this information they predict the rate of inflation as well as they can. It is in this sense that the expectations are 'rational': people use their reason to assess the future on the basis of current information.

But forecasters frequently get it wrong, and so do economic commentators! And the government does not always do what it says it will. Thus workers and firms will be basing expectations on *imperfect information*. The crucial point about the rational expectations theory, however, is that these errors in prediction are *random*. People's predictions of inflation are just as likely to be too high as too low.

Anticipating fluctuations in aggregate demand

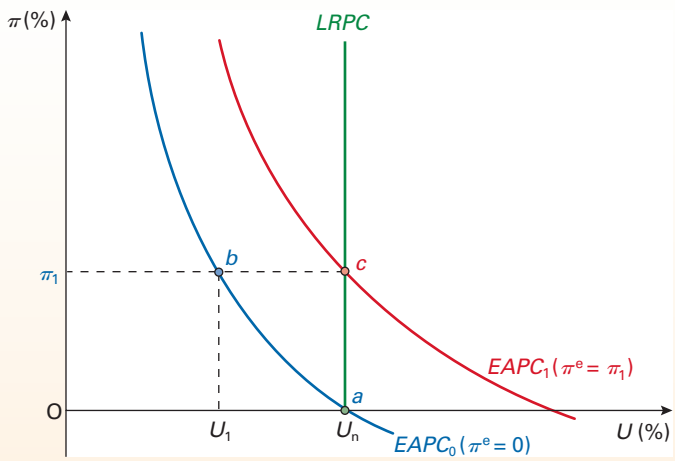
Assume that the government raises aggregate demand and that the increase is expected. People will anticipate that this will lead to higher prices and wages. If both goods and labour markets clear continuously because of the flexibility of prices and wages, there will be *no* effect on output and employment. If their expectations of higher inflation are correct, this will *fully* absorb the increase in nominal aggregate demand, such that there will have been no increase in *real* aggregate demand at all. Firms will not produce any more output or employ any more people: after all, why should they? If they anticipate that people will spend 10 per cent more money, but that prices will rise by 10 per cent, their *volume* of sales will remain the same.

We can use Figure 12.16 to illustrate the adjustment of the economy under continuous market clearing, rational expectations and anticipated demand shocks. Assume that the economy is at point *a* with unemployment at its natural rate, U_n, and an actual and expected inflation rate of zero.

Now assume that government increases aggregate demand. With rational expectations and no surprises, people fully anticipate that the inflation rate will rise to π_1. $EAPC_0$ which is based on expectations of zero inflation cannot be moved along. The moment that aggregate demand rises people correctly anticipate an inflation rate of π_1. Thus the whole *EAPC* moves vertically upwards to $EAPC_1$. As a result, the economy moves *directly* from point *a* to point *c*.

Output and employment will only rise, therefore, if people make an error in their predictions (i.e. if they under-predict the rate of inflation and interpret an increase in money spent as an increase in *real* demand).

Figure 12.16 Anticipated and unanticipated changes in aggregate demand



But they are as likely to *over*-predict the rate of inflation, in which case output and employment will fall! Thus there is no *systematic* trade-off between inflation and unemployment, even in the short run.

The vertical short-run Phillips curve is therefore comparable to the vertical short-run aggregate supply curve we saw in Section 10.2 (see pages 280–1). The vertical *SRAS* curve shows aggregate supply (output) being determined independently of aggregate demand.

Both the vertical short-run Phillips curve and the vertical *SRAS* curve can be used to illustrate how anticipated changes in economic policy, such as changes in government spending, have no effect on output and employment. Instead, they remain at their equilibrium levels. This controversial conclusion is known as the **policy ineffectiveness proposition**.

Pause for thought

For what reasons would a new classical economist support the policy of a central bank publishing its inflation forecasts and the minutes of its deliberations about interest rates?

Keynesian views

Keynesians criticise the monetarist/new classical approach of focusing exclusively on price expectations. Expectations, argue Keynesians, influence *output* and *employment* decisions, not just pricing decisions.

TC 11 p 61

Definition

Policy ineffectiveness proposition The conclusion drawn from new classical models that, when economic agents anticipate changes in economic policy, output and employment remain at their equilibrium (or natural) levels.

If there is a gradual but sustained expansion of aggregate demand, firms, seeing the economy expanding and seeing their orders growing, will start to invest more and make longer-term plans for expanding their labour force. People will generally *expect* a higher level of output, and this optimism will cause that higher level of output to be produced, both from existing capacity and from new capacity arising from the increased investment. In other words, expectations will affect output and employment as well as prices, not only in the short term, but over the longer term too.

Graphically, as was shown in Figure 10.7 (page 282), the increased output and employment from the recovery in investment will shift the *SRAS* curve to the right. It will also shift the Phillips curve to the left, offsetting (partially, wholly or more than wholly) the upward shift from higher inflationary expectations.

The lesson here for governments is that a sustained, but moderate, increase in aggregate demand can lead to a sustained growth in aggregate supply. What should be avoided is an excessive and unsustainable expansion of aggregate demand.

As the last two sections have shown, there are contrasting perspectives on the relationship between output, unemployment and inflation and how they are affected by changes in aggregate demand. The extraordinary impact of the global financial crisis and the COVID-19 pandemic on aggregate levels of spending helped to fuel debates around whether significant changes in aggregate demand might lead to enduring effects on output and unemployment. This raises questions about the scale and nature of government interventions in such cases. Chapter 13 looks at government policy in more detail.

Pause for thought

Why is it important in the Keynesian analysis for there to be a steady expansion of aggregate demand?

Recap

1. Expectations can be incorporated into the analysis of the Phillips curve. The effect is to generate a series of expectations augmented Phillips curves (*EAPC*s).
2. The natural rate hypothesis assumes that the natural or equilibrium rate of unemployment is determined independently of aggregate demand. The result is a vertical long-run curve at the equilibrium rate of unemployment.
3. Adaptive expectations are backward-looking expectations. The simplest form of adaptive expectations of inflation assumes that the expected rate of inflation this year is what it actually was last year: $\pi^e_t = \pi_{t-1}$.
4. The monetarist model assumes that markets are relatively flexible. But the assumption of adaptive expectations allows unemployment to deviate from the natural rate in the short term. People can be fooled, allowing unemployment and output to fluctuate.
5. If there is excess demand in the economy, producing upward pressure on wages and prices, initially unemployment will fall. The reason is that workers and firms will believe that wage and price increases represent *real* wage and price increases respectively. Thus workers are prepared to take jobs more readily and firms choose to produce more. But as people's expectations adapt upwards to these higher wages and prices, so ever-increasing rises in nominal aggregate demand will be necessary to maintain unemployment below the natural rate. Price and wage rises will accelerate: i.e. inflation will rise.

6. The long-run Phillips curve, according to this analysis, is thus vertical at the natural rate of unemployment.
7. New classical theories assume continuous market clearing with flexible prices and wages in the short run as well as in the long run. It also assumes that people base their expectations of inflation on a rational assessment of the *current* situation.
8. People may predict wrongly, but they are equally likely to underpredict or to overpredict. On average, over the years, they will predict correctly.
9. The assumptions of continuous market clearing and rational expectations imply that only unexpected fluctuations in aggregate demand will cause unemployment to deviate from its natural rate. There can be no short-run trade-off between inflation and unemployment when changes in aggregate demand are anticipated.
10. If people correctly predict the rate of inflation, they will correctly predict that any increase in nominal aggregate demand will simply be reflected in higher prices. Total output and employment will remain the same: at the natural level.
11. If people underpredict the rate of inflation, they will believe that there has been a *real* increase in aggregate demand, and thus output and employment will increase. But they are just as likely to overpredict the rate of inflation, in which case they will believe that real aggregate demand has fallen. The result is that output and employment will fall.
12. Keynesians argue that if people expect a rise in aggregate demand to be sustained, firms will invest more, thereby increasing potential output (and possibly also reducing unemployment) and not just increasing the rate of inflation. Conversely, austerity policies can reduce potential as well as actual output.

12.5 INFLATION RATE TARGETING

Does successful inflation rate targeting produce a horizontal Phillips curve?

The period from the early 1990s up to the financial crisis of the late 2000s was to be known as the 'Great Moderation'. The period was characterised in many developed counties by low and stable inflation and continuous economic growth. In the UK, for example, between 1994 to 2007 annual UK growth averaged 3 per cent, while the annual rate of consumer price inflation averaged 1.75 per cent.

The period of the Great Moderation also saw the emergence of a new mainstream macroeconomic consensus. The consensus was a fusion of ideas drawn from different macroeconomic perspectives. It had two central elements.

- First, there is no long-run trade-off between inflation and unemployment (the natural rate hypothesis). The natural rate of unemployment and potential output are argued to be principally determined by supply-side or structural factors.
- Second, fluctuations in aggregate demand and supply can lead to short-term deviations of output from its potential level and unemployment from its natural rate. The extent of these deviations depends on market imperfections, including the inflexibility of goods prices and wage rates.

In the UK, as in many other industrialised countries, the Great Moderation was a period when policy makers emphasised the importance of maintaining a stable economic environment and focused on raising long-term economic growth. This led policy makers to apply **constrained discretion**: a set of rules or principles providing a framework for economic policy.

Constrained discretion typically involves the use of targets, such as an inflation target or a public-sector deficit or debt ceiling. The key is to affect the *expectations* of the public, for example in relation to inflation.

TC 11 p 61

Delegating monetary policy

In many countries, governments handed over the operation of monetary policy to central banks. This is known as the **delegation of monetary policy**. In the UK, in 1997 the Bank of England was granted independence to determine interest rates to meet an inflation rate target.

The natural rate hypothesis provides the theoretical grounds for central bank independence. If there is no long-run trade-off between inflation and unemployment then, monetarists and new classical macroeconomists argue, it makes sense to remove the temptation for governments to use expansionary monetary policy unexpectedly, and perhaps simply for their own political benefit.

In the absence of delegating monetary policy and operating a transparent inflation rate target, the result

Definitions

Constrained discretion A set of principles or rules within which economic policy operates. These can be informal or enshrined in law.

Delegation of monetary policy The handing over by government of the operation of monetary policy to central banks.

is likely to be higher inflation rates. This is because the public will tend to form higher expectations of inflation, which lead to higher actual inflation rates. The public do this since they are aware of governments' incentive to want to reduce unemployment (and increase output) by loosening monetary policy and creating additional, unexpected inflation. Higher expectations of inflation choke off increases in *real* aggregate demand, output and employment that the government might try to bring about.

TC 11 p 61

TC 5 p 17

The longer-term position, it is argued, is therefore one of higher rates of inflation, yet with output and unemployment at their natural levels. There is no additional output or employment, merely additional and hence excessive inflation. This excessive inflation is known as **inflation bias**.

Pause for thought

Why would delegating monetary policy reduce inflation bias?

The elimination of inflation bias is commonly identified as the key economic benefit of central bank independence. The delegation of monetary policy to the central bank with a clear mandate that it then sticks to is said to enhance the **credibility of monetary policy**.

By sticking to an inflation target, a central bank can create the stable environment necessary for the market to flourish: expectations will adjust to the target rate of inflation (assuming central banks are successful in achieving the target) and firms will be able to plan with more confidence. Investment is thereby encouraged and this, in turn, encourages a growth in potential output. In other words, sticking to the targets creates the best environment for the expansion of aggregate *supply*.

Inflation targeting

For many countries inflation targeting has become central to their macroeconomic policy. Table 12.3 gives the targets for a selection of countries (as of 2022).

Definitions

Inflation bias Excessive inflation that results from people raising their expectations of the inflation rate following expansionary demand management policy, encouraging government to loosen policy even further.

Credibility of monetary policy The extent to which the public believes that the central bank will take the measures necessary to achieve the stated targets of monetary policy, for example an inflation rate target.

Table 12.3 Inflation rate targets of central banks (2018)

Country	Inflation target (%)	Details
Australia	2–3	Average over the medium term
Brazil	3.5	Tolerance band of ±1.5 percentage points
Canada	2	Tolerance band of ±1 percentage point
Chile	3	Tolerance band of ±1 percentage point
Czech Republic	2	Tolerance band of ±1 percentage point
Eurozone	2	Average for eurozone as a whole; over medium term
Hungary	3	Tolerance band of ±1 percentage point
Iceland	2.5	
Israel	1–3	
Japan	2	
Mexico	3	Tolerance band of ±1 percentage point
New Zealand	2	Tolerance band of ±1 percentage point
Norway	2	Close to 2 per cent over time
Peru	2	Tolerance band of ±1 percentage point
Poland	2.5	Tolerance band of ±1 percentage point
South Africa	3–6	
South Korea	2	
Sweden	2	
Switzerland	<2	
Thailand	1–3	Tolerance band of ±1.5 percentage points
UK	2	Forward-looking inflation target; tolerance band of ±1 percentage point
USA	2	Average over time; consistent with dual mandate of stable prices and maximum employment

Source: Various central bank websites (see BIS *Central Bank Hub*; see also *Central Bank News*)

The success of inflation-rate targeting depends crucially on its impact on inflationary expectations. Many countries up to the inflation shocks of 2021–22 had been relatively successful in delivering low and stable inflation rates. Indeed, the more successful they were, the more people expected this success to be maintained, which in turn helped to ensure this success. Take the UK, for example. The average rate of CPI inflation from 1997 to 2021 was 2 per cent – the Bank of England's inflation-rate target.

So, have there been any problems with inflation targeting?

The horizontal Phillips curve. Ironically, one of the main problems may have been in its success. With worldwide inflation having fallen from the 1990s, and with global trade and competition helping to keep prices down, there became less of a link between inflation and the business cycle. The Phillips curve became close to horizontal (see Figure 12.14). Thus, economies saw movements left or right *along* this horizontal line from one year to the next depending on the level of economic activity. Such fluctuations in unemployment had become consistent with a stable inflation rate.

Under inflation rate targeting, booms may no longer generate the inflation they once did. Therefore, gearing interest rate policy to maintaining low inflation could still see economies experiencing unsustainable booms, followed by recessions. Inflation may be controlled, but the business cycle may not be.

TC 3 p 10

A further complication is that there may be movements left or right along the horizontal Phillips curve depending on what happens to equilibrium unemployment. A reduction in equilibrium unemployment will result in a leftward movement, while an increase will lead to a rightward one.

Economic and financial cycles. When the focus of monetary policy is largely on inflation, this can result in greater fluctuations in output. After all, it is not the direct impact of economic shocks on economic activity that is necessarily of concern but rather their indirect impact on inflation.

Also, the low interest rates seen in many countries during the first half of the 2000s helped to fuel unsustainable flows of lending. While low interest rates were consistent with meeting the inflation rate target, they encouraged property purchase and other investments. These stretched the financial well-being of economic agents when interest rates rose. A point was reached when a significant retrenchment by banks, businesses and individuals became inevitable. This is the idea of a balance sheet recession (see page 265).

KI 32 p 239

Cost-push inflation. When central banks use interest rates to manage demand-pull inflation, there is what economists call a **divine coincidence**: the reduction or absence of a trade-off between stabilising inflation and stabilising output around potential output. If demand-pull inflation is increasing because the economy is nearing or exceeding its potential output, then raising interest rates is appropriate both to manage the output gap and to reduce demand-pull inflation. Equally, cutting interest rates is appropriate if the rate of inflation is falling and a negative output gap is emerging.

However, when cost-push inflation is the principal cause of inflation then there becomes a trade-off between managing inflation and the output gap. This is because cost-push inflation raises prices but reduces output (see page 355). A rise in interest rates in such circumstances will help to reduce inflation, but will worsen the negative output gap.

This type of policy dilemma is one that central banks have faced increasingly in recent times. There are two significant examples of this. First, there were the inflationary problems from 2010 to 2012, just after the financial crisis. Several countries, including the UK, had only just emerged from recession but commodity prices were rising. Most central banks chose to keep interest rates at historic lows.

Second, there were the inflationary shocks of 2021–22. Once more, rising commodity prices, along with other supply-side factors, such as rising transportation costs and shortages of important inputs, were helping to raise inflation rates. Yet economies were just emerging from historically-large contractions resulting from the restrictions imposed to tackle the COVID-19 pandemic. This time, central banks did raise rates, though tentatively at first, as concerns grew that high inflation rates could persist if action were not taken. But these rises in interest rates helped to drive several economies into recession.

Pause for thought

What policy dilemma might cost-push deflation *create for a central bank?*

Chapter 13 looks in more detail at the various types of policy the government and the central bank can use to tackle the various macroeconomic problems and issues that economies face.

Definition

Divine coincidence (**in monetary policy**) When the monetary authorities can choose a policy stance that is largely consistent with both stabilising inflation and stabilising output around potential output.

Recap

1. The delegation by government of the operation of monetary policy to central banks and the adoption of inflation rate targeting is argued to reduce inflation bias. It enhances the credibility of monetary policy which enables low and stable inflation.
2. Inflation targeting in the UK has seen the rate of inflation generally close to the target level. Inflation rate targeting helps to anchor inflationary expectations around the target rate. This has tended to make the time-path of the Phillips curve horizontal at the target rate of inflation.
3. Some economists argue that inflation rate targeting is not without issues. These include: stages of the business cycle not readily identifiable by the inflation rate; monetary policy not accounting for other macroeconomic concerns such as the sustainability of bank lending; and the policy dilemma for central banks when cost-push inflationary pressures develop, such as those following the financial crisis and COVID-19 pandemic.

QUESTIONS

1. Would it be desirable to have zero unemployment?
2. What major structural changes have taken place in the UK economy in the past couple of decades that have increased or decreased structural unemployment?
3. What are the causes of unemployment in the area where you live?
4. What would be the benefits and costs of increasing the rate of unemployment benefit?
5. Consider the most appropriate policies for tackling each of the different types of unemployment.
6. Do any groups of people gain from inflation?
7. If everyone's incomes rose in line with inflation, would it matter if inflation were 100 per cent or even 1000 per cent per annum?
8. Imagine that you had to determine whether a particular period of inflation was demand-pull, or cost-push, or a combination of the two. What information would you require in order to conduct your analysis?
9. What is the difference between adaptive explanations and rational expectations?
10. Assume that there is a trade-off between unemployment and inflation, traced out by a 'Phillips curve'. What could cause a leftward shift in this curve?
11. Assume that inflation depends on two things: the level of aggregate demand, indicated by the inverse of unemployment ($1/U$), and the expected rate of inflation (π_t^e). Assume that the rate of inflation (π_t) is given by the equation:

    $$\pi_t = (48/U - 6) + \pi_t^e$$

    Assume initially (year 0) that the actual and expected rate of inflation is zero.

    a. Now assume in year 1 that the government wishes to reduce unemployment to 4 per cent and continues to expand aggregate demand by as much as is necessary to achieve this. Fill in the rows for years 0 to 4 in the following table. It is assumed for simplicity that the expected rate of inflation in a given year (π_t^e) is equal to the actual rate of inflation in the previous year (π_{t-1})

    b. Now assume in year 5 that the government, worried about rising inflation, reduces aggregate demand sufficiently to reduce inflation by 3 per cent in that year. What must the rate of unemployment be raised to in that year?

    c. Assuming that unemployment stays at this high level, continue the table for years 5 to 7.

Year	U	$48/U - 6$	+	π^e	=	π
0	...	...	+	...	=	...
1	...	...	+	...	=	...
2	...	...	+	...	=	...
3	...	...	+	...	=	...
4	...	...	+	...	=	...
5	...	...	+	...	=	....
6	...	...	+	...	=	...
7	...	...	+	...	=	...

12. In the accelerationist model, if the government tries to maintain unemployment below the natural rate, what will determine the speed at which inflation accelerates?
13. How might fluctuations in aggregate demand lead to changes in output even under the assumption of continuous market clearing?
14. For what reasons may the equilibrium rate of unemployment increase?
15. Given the persistence of high levels of unemployment after the recession of the early 1980s, what policies would you advocate to reduce unemployment in such circumstances?
16. Explain each of the following:
    a. Why there were simultaneously higher rates of inflation *and* unemployment in the 1970s and 1980s than in the 1950s and 1960s.
    b. Why there were simultaneously lower rates of inflation *and* unemployment in the late 1990s and early 2000s than in the 1970s and 1980s?
17. In what sense is it true to say that the Phillips curve has become horizontal?
18. What is meant by inflation bias? What factors might affect the potential magnitude of inflation bias?
19. What is meant by the credibility of monetary policy? How could its credibility be assessed?

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

12.1 **Swimming in the unemployment pool.** The composition and duration of unemployment in the EU.

12.2 **Classical 'remedies' for unemployment.** How the policies advocated by the classical economists to cure unemployment would, according to Keynes, make the problem worse.

12.3 **Technology and unemployment.** Does technological progress destroy jobs?

12.4 **Disinflation in Europe and Japan.** What happens when there is negative inflation? Is it a case of survival of the fittest?

12.5 **CPI, CPIH and RPI inflation.** This examines the different measures of consumer price inflation published in the UK.

12.6 **Hyperinflation.** This looks at the extraordinarily high rates of inflation experienced in Germany in the early 1920s, Serbia and Montenegro in the 1990s and in Zimbabwe in the 2000s.

12.7 **Money and inflation in ancient Rome.** A very early case study of the quantity theory of money: how the minting of extra coins by the Romans caused prices to rise.

12.8 **A.W.H. Phillips.** A portrait of the shy economist and engineer who invented the famous Phillips curve and Phillips machine.

12.9 **Explaining the shape of the short-run Phillips curve.** This shows how money illusion on the part of workers can explain why the Phillips curve is downward sloping.

12.10 **Cost-push factors in Keynesian analysis.** How Keynesians incorporated cost-push inflation into their analysis of shifts in the Phillips curve.

12.11 **Milton Friedman (1912–2008).** A profile of the most influential of the monetarist economists.

12.12 **The political business cycle.** This shows how the accelerationist theory and 'clockwise Phillips loops' can be used to analyse the so-called political business cycle.

12.13 **Getting predictions wrong.** How incorrect predictions can lead to a rise or fall in output in the new classical model.

12.14 **The rational expectations revolution.** A profile of two of the most famous economists of the new classical rational expectations school.

12.15 **A spectrum of views.** An overview of the different schools of macroeconomic thought.

12.16 **Common ground between economists?** A discussion of the 'new consensus' among macroeconomists that is argued to have developed through the 1990s and into the 2000s.

Chapter 13

Macroeconomic policy

This chapter focuses on the three main types of macroeconomic policy: fiscal, monetary and supply-side policy.

Fiscal policy seeks to control aggregate demand by altering the balance between government expenditures and taxation. Monetary policy seeks to control aggregate demand by directly controlling money supply or by altering the rate of interest and then backing this up by any necessary changes in money supply.

Both fiscal and monetary policy can be used to control aggregate demand. Excessive growth in aggregate demand can cause unsustainable short-term growth and higher rates of inflation. Too little aggregate demand can result in a recession, with negative growth and rising unemployment.

The first three sections of this chapter examine how both fiscal and monetary policy work; how effective they are likely to be in controlling aggregate demand; and what are the potential pitfalls in their use are.

We then turn to consider the arguments as to how much discretion or control policy makers should have over fiscal and monetary policy. Should governments adopt fixed targets for policy (e.g. inflation targets) or should they adjust policies according to circumstances? This is a debate reinvigorated by the global financial crisis of 2007–9, the Russian invasion of Ukraine and the COVID-19 pandemic and by their impact on economies and wider society.

In the final section we turn to the supply side. We will look at various policies that aim to increase aggregate supply.

Supply-side policies can be put into two broad categories: market orientated and interventionist. Market-orientated policies focus on 'freeing up' markets and improving market incentives. They involve policies such as tax cuts, privatisation and deregulation. Interventionist policies, by contrast, focus on ways of countering the inadequacies of markets through direct government provision of transport infrastructure, training or R&D, or financial support for private provision.

After studying this chapter, you should be able to answer the following questions:

- What are the different types of macroeconomic policy that governments or central banks can use to control the economy?
- What are the principal measures of a country's public finances?
- What is meant by the sustainability of the public finances?
- How can fiscal policy be used to alter the level of aggregate demand? What are the strengths and drawbacks of using fiscal policy to (a) stabilise the economy; (b) cure fundamental disequilibria?
- How is monetary policy operated by the central bank? How does it alter the money supply? How does it control interest rates? What special measures have been introduced since the financial crisis to stimulate the economy?
- How successful is monetary policy likely to be in controlling (a) aggregate demand, and (b) inflation?
- Should fiscal and monetary policy be frequently adjusted in an attempt to stabilise the economy? Or should the government set fiscal and monetary rules, which are then rigidly followed?
- What types of supply-side policy are available to governments? What are their strengths and weaknesses?

13.1 FISCAL POLICY AND THE PUBLIC FINANCES

What are some of the key indicators of the public finances?

Fiscal policy involves the government manipulating the level of government expenditure and/or rates of tax. An *expansionary* fiscal policy will involve raising government expenditure (an injection into the circular flow of income) or reducing taxes (a withdrawal from the circular flow). A *deflationary* (i.e. a contractionary) fiscal policy will involve cutting government expenditure and/or raising taxes.

TC 14 p 249

Roles for fiscal policy

Fiscal policy may be used to affect the level of aggregate demand. There are two principal reasons for this.

Prevent the occurrence of fundamental disequilibrium in the economy. The government may wish to remove any severe deflationary or inflationary gaps. Hence, expansionary fiscal policy could be used to prevent an economy experiencing a severe or prolonged recession, such as that experienced in the Great Depression of the 1930s, the global financial crisis of 2007–9 and the COVID-19 pandemic of the early 2020s, when substantial tax cuts and increased government expenditure were used by many countries to combat the onset of recession. Likewise, deflationary fiscal policy could be used to prevent rampant inflation, such as that experienced in the 1970s.

TC 14 p 249

Stabilisation policies. The government may wish to smooth out the fluctuations in the economy associated with the business cycle. This involves reducing government expenditure or raising taxes when the economy begins to boom. This will dampen down the expansion and prevent 'overheating' of the economy, with its attendant problems of rising inflation and a deteriorating current account balance of payments. Conversely if a recession looms, the government should cut taxes or raise government expenditure in order to boost the economy.

TC 7 p 26

Fiscal policy can also be used to influence aggregate *supply* – to increase the growth of potential output and reduce the natural rate of unemployment. During the COVID-19 pandemic, policies were adopted to protect jobs and businesses. Part of the reason was to allow economies to recover more seamlessly when health intervention measures were able to be relaxed.

Public-sector finances

Government finances

The term 'government' is often used interchangeably with that of 'general government' when analysing government finances. General government includes both *central* and *local* government.

The terms **budget deficit** and **budget surplus** are frequently used in the context of government, and especially *central* government. However, these terms can be applied to any organisation to assess its financial well-being by comparing expenditure with revenues.

KI 23 p 173

If the total expenditure (including benefits) of both central and local government exceeds the revenue from taxation, council taxes, rates, etc., this is known

Definitions

Fiscal policy Policy to affect aggregate demand by altering government expenditure and/or taxation.

Budget deficit The excess of an organisation's spending over its revenues. When applied to government it is the excess of its spending over its tax receipts.

Budget surplus The excess of an organisation's revenues over its expenditures. When applied to government it is the excess of its tax receipts over its spending.

as a **general government deficit**. Conversely, when general government's revenues exceed its expenditure there is a **general government surplus.**

For most of the past 50 years governments around the world have run deficits. The situation improved in many countries from the late 1990s due to a mix of strong economic growth and government attempts to reduce deficits. The position changed dramatically, however, following first the global financial crisis and later the COVID-19 pandemic as governments around the world increased their expenditure and cut taxes in an attempt to stave off recession. Government deficits soared.

KI 23 p 173

Deficits, debt and borrowing. To finance their deficits, governments will have to borrow (e.g. through the issue of bonds (gilts) or Treasury bills). As we saw in Section 11.3, this will lead to an increase in the money supply to the extent that the borrowing is from the banking sector. The purchase of bonds or Treasury bills by the (non-bank) private sector, however, will not lead to an increase in the money supply.

Deficits represent *annual* borrowing: a flow concept. The accumulated deficits over the years (minus any surpluses) gives total *debt*: a stock concept. It is the total amount owed by the government. Central and general government debt are known as **national debt** and **general government debt**, respectively.

KI 23 p 173

Note that the national debt is not the same thing as the country's overseas debt. In the case of the UK, only around one quarter of gilts are held overseas. The remainder is held by UK financial institutions and some individual UK residents. In other words, the government finances its budget deficits largely by borrowing at home and not from abroad.

Table 13.1 shows general government deficits/surpluses and debt for selected countries. They are expressed as a proportion of GDP.

Pause for thought

Why are historical and international comparisons of deficit and debt measures best presented as proportions of GDP?

Definitions

General government deficit (or surplus) The combined deficit (or surplus) of central and local government.

National debt The accumulated deficits of central government. It is the total amount owed by central government, both domestically and internationally.

General government debt The accumulated deficits of central plus local government. It is the total amount owed by general government, both domestically and internationally.

Table 13.1 General government deficits/surpluses and debt as a percentage of GDP

	General government deficits (−) or surpluses (+)			General government debt		
	Average 1995–2007	Average 2008–19	Average 2020–23	Average 1995–2007	Average 2008–19	Average 2020–23
Austria	−2.8	−1.9	−4.6	66.3	79.3	80.9
Belgium	−1.3	−2.8	−6.0	108.9	102.0	109.0
France	−3.0	−4.3	−5.8	61.8	90.9	112.0
Germany	−3.1	−0.1	−2.9	61.2	71.7	67.2
Greece	−6.7	−6.4	−5.7	102.4	165.3	191.4
Ireland	1.3	−7.3	−1.8	41.6	82.0	52.6
Italy	−3.5	−3.0	−6.7	110.4	127.4	150.2
Japan	−5.7	−5.8	−9.7	140.6	221.2	262.5
Netherlands	−1.5	−1.8	−2.8	55.4	59.9	52.2
Norway	10.1	9.7	5.0	39.8	36.6	41.2
Portugal	−4.4	−5.1	−2.9	62.8	116.5	124.5
Spain	−1.4	−6.6	−6.6	52.9	85.5	116.8
Sweden	−0.1	0.0	−0.7	54.7	39.7	35.2
UK	−2.0	−5.5	−6.8	39.2	78.5	101.0
USA	−3.3	−7.6	−9.4	61.5	98.9	126.8
Eurozone	−2.7	−2.8	−4.6	70.1	88.1	96.0

Note: Data from 2022 are forecasts
Source: Based on data from *AMECO Database*, Tables 16.3 and 18.1 (European Commission, DG ECFIN)

As you can see, in the period from 1997 to 2007, all the countries, with the exception of Ireland and Norway, ran an average deficit. In the period from 2008 to 2019 the average deficits increased for most countries before increasing again from 2020 as governments responded to the COVID-19 pandemic. And the bigger the deficit, the faster debt increases.

Public sector

To get a more complete view of public finances, we would need to look at the spending and receipts of the entire public sector: namely, central government, local government and public corporations.

Total spending and receipts. Figure 13.1 shows UK public-sector spending and receipts as a proportion of GDP. After significant rises after both world wars, their size as a share of GDP then tended to fluctuate around the 40 per cent mark, depending on the effect of the business cycle on spending and receipts. The financial crisis and the COVID-19 pandemic saw UK public spending temporarily rising to 46 and 52 per cent of GDP respectively.

Current and capital expenditures. In presenting the public finances, it has become custom to distinguish between **current** and **capital expenditures**. Current expenditures include items such as wages and salaries of public-sector staff, administration and the payments of welfare benefits. Capital expenditures give rise to a stream of benefits *over time*. Examples include expenditure on roads, hospitals and schools. In recent years, gross capital spending has typically been around 10 per cent of UK public-sector spending.

Pause for thought

What could have driven the changes in the composition of UK public expenditure? Do such changes matter?

Final expenditure and transfers. We can also distinguish between **final expenditure** on goods and services, and **transfers**. This distinction recognises that the public sector directly adds to the economy's aggregate demand (an injection into the circular flow of income: see page 245) through its spending on goods and services, including the wages of public-sector workers, but also that it redistributes incomes

Definitions

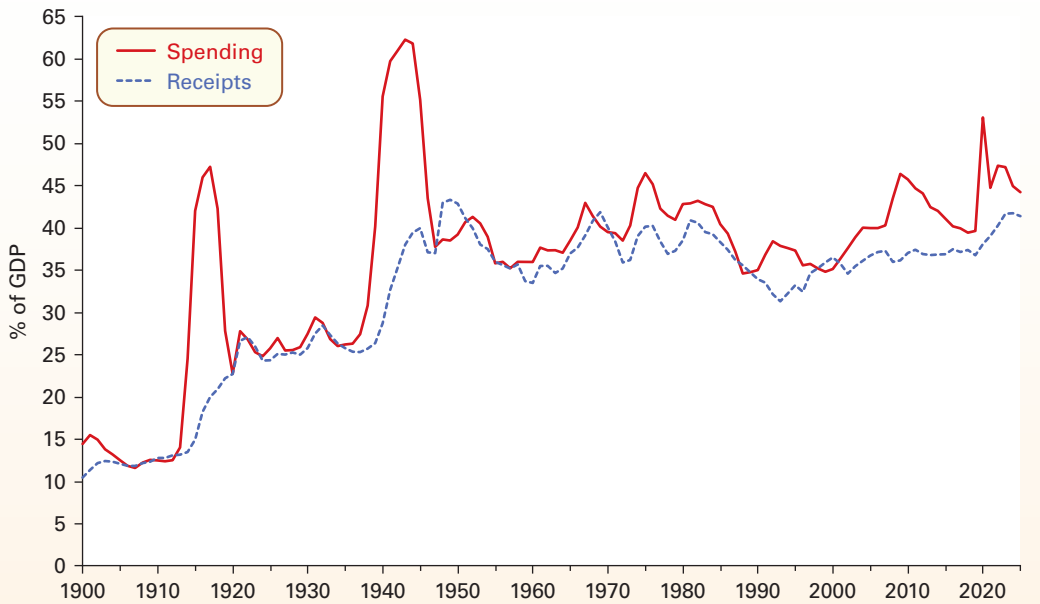
Current expenditure Recurrent spending on goods and factor payments.

Capital expenditure Investment expenditure; expenditure on assets.

Final expenditure Expenditure on goods and services. This is included in GDP and is part of aggregate demand.

Transfers Transfers of money from taxpayers to recipients of benefits and subsidies. They are not an injection into the circular flow but are the equivalent of a negative tax (i.e. a negative withdrawal).

Figure 13.1 UK public-sector spending and receipts (financial years)



Notes: OBR forecasts from 2022/23; figures exclude publicly-owned banks.

Sources: *Public Finances Databank*, Office for Budget Responsibility (November 2022)

between individuals and firms. Transfers include subsidies and benefit payments, such as payments to the unemployed. These are not injections into the circular flow but count, instead, as negative taxes as they increase households' disposable income

Public-sector deficits. If the public sector spends more than it earns, it will have to finance the deficit through borrowing: known as the **public-sector net borrowing** (**PSNB**). The principal form of borrowing is through the sale of gilts (bonds). The precise amount of money borrowed in any one year is known as the **public-sector net cash requirement** (**PSNCR**). It differs slightly from the PSNB because of time lags in the flows of public-sector incomes and expenditure.

KI 23 p 173

Public-sector (gross) debt. As with central and general government debt, public-sector debt is the current stock of the accumulated deficits over the years. Another measure is **public-sector net debt**. This is the sector's gross debt *less* its liquid assets, which comprise official reserves and deposits held with financial institutions.

Sustainability of public finances

KI 23 p 173

The sustainability of the public finances can be assessed by considering the ability of the public sector to maintain or reduce its debt-to-GDP ratio. Should the ratio continue rising, the public sector faces increasing financial pressure to meets it current debt obligations and diverts more resources to these and away from other spending options.

Crucial to whether the public-sector debt-to-GDP ratio rises or falls are the current value of the ratio, the expected rate of interest on outstanding debts, the expected rate of economic growth and, finally, the public sector's **primary surplus** (**or deficit**). A primary surplus (deficit) occurs when public-sector receipts are greater (less) than public-sector expenditures *excluding* interest payments.

In fact, for the public sector to be able to run a primary deficit without the debt-to-GDP ratio rising, the rate of economic growth must be greater than the rate of interest. Therefore, the sustainability constraints on spending and/or the need to raise revenue from taxation are greater the higher the debt-to-GDP ratio and the lower the rate of economic growth are.

Box 13.1 considers the so-called 'fiscal arithmetic' behind the sustainability of the public finances in more detail.

The business cycle and the public finances

The size of the deficit or surplus is not entirely due to deliberate government policy. It is also influenced by the state of the economy. Therefore, when analysing the finances of government or the whole public sector, it is important to recognise that there is both a cyclical and a structural component.

If the economy is booming with people earning high incomes, the amount paid in taxes will be high. Also, with low unemployment, less will be paid in unemployment benefits. The combined effect of increased tax revenues and reduced benefits is to reduce the public-sector deficit (or increase the surplus).

By contrast, if the economy is depressed, tax revenues will be low and the amount paid in benefits will be high. This will increase the public-sector deficit (or reduce the surplus).

By 'cyclically adjusting' measures of public-sector deficits or surpluses we remove their cyclical component. In other words, we show just the direct effects of government policy, not the effects of the level of economic activity.

Figure 13.2 shows both actual and cyclically-adjusted public-sector net borrowing in the UK as a percentage of GDP since the mid 1970s. Over the long run the economy's output gap is zero (see Box 9.1). Hence, over the period shown, both net borrowing measures average the same (around 3.5 per cent of GDP).

The deficit or surplus that would arise if the economy were producing at the potential level of national income is termed the **structural deficit or surplus**. Remember that the potential level of national income is where there is no excess or deficiency of aggregate demand: where there is a zero-output gap.

The fiscal stance

The government's **fiscal stance** refers to whether it is pursuing an expansionary or contractionary fiscal policy. Does the fact that countries such as the UK in

Definitions

Public-sector net borrowing (**PSNB**) The difference between the expenditures of the public sector and its receipts from taxation and the revenues from public corporations.

Public-sector net cash requirement (**PSNCR**) The (annual) deficit of the public sector, and thus the amount that the public sector must borrow. In the UK the principal measure, which takes into account financial transactions by the public sector, is known as the public-sector net cash requirement.

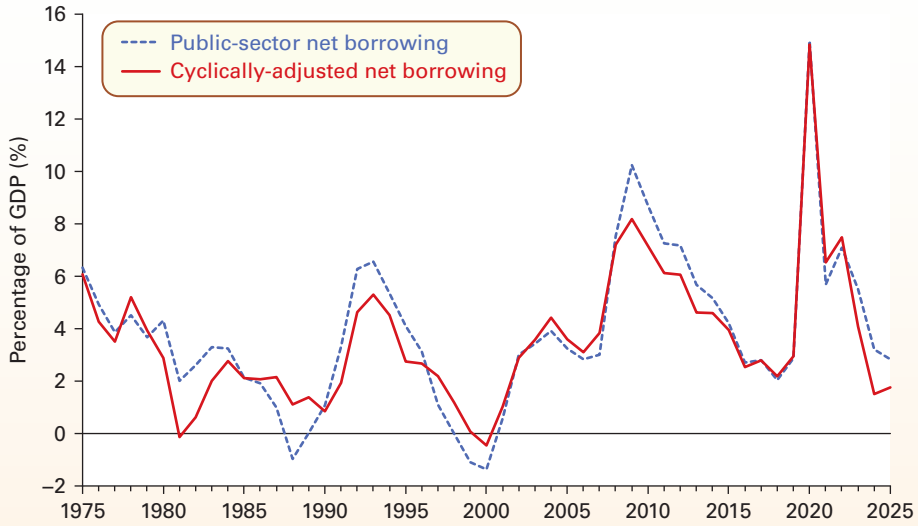
Public-sector net debt Gross public-sector debt minus liquid financial assets.

Primary surplus (**or deficit**) Public-sector receipts minus public-sector expenditure *excluding* interest payments.

Structural deficit (**or surplus**) The public-sector deficit (or surplus) that would occur if the economy were operating at the potential level of national income: i.e. one where there is a zero-output gap.

Fiscal stance How expansionary or contractionary the Budget is.

Figure 13.2 UK public-sector net borrowing



Notes: OBR forecasts from 2022/23; figures exclude public banks.

Source: *Public Finances Databank* (Office for Budget Responsibility, November 2022)

most years run public-sector deficits mean that their government's fiscal stance is mainly expansionary? Would the mere existence of a surplus mean that the stance was contractionary? The answer is no. Whether the economy expands or contracts depends on the balance of *total* injections and *total* withdrawals.

What we need to focus on is *changes* in the size of the deficit or surplus. If the deficit this year is lower than last year, then (other things being equal) aggregate demand will be lower this year than last. The reason is that either government expenditure (an injection) must have fallen, or tax revenues (a withdrawal) must have increased, or a combination of the two.

To conclude, the size of the deficit or surplus is a poor guide to the stance of fiscal policy. A large deficit *may* be due to a deliberate policy of increasing aggregate demand, but it may be due simply to the fact that the economy is depressed.

Recap

1. The public sector comprises general government and public corporations. There exist a range of fiscal indicators that allow us to analyse the fiscal position and well-being of the public sector or its component sectors.
2. The sustainability of a country's public finances can be analysed by the fiscal arithmetic needed to maintain the current public-sector debt-to-GDP ratio. Countries with higher existing debt-to-GDP ratios, higher real interest costs and lower real rates of economic growth will need to operate larger primary surpluses (or smaller deficits) to prevent the ratio from rising further.
3. The government's fiscal policy influences the size of its budget deficit or surplus. Its size alone, however, is a poor guide to the government's fiscal stance. A large deficit, for example, may simply be due to the fact that the economy is in recession and therefore tax receipts are low. A better guide is whether the change in the deficit or surplus will be expansionary or contractionary.

13.2 THE USE OF FISCAL POLICY

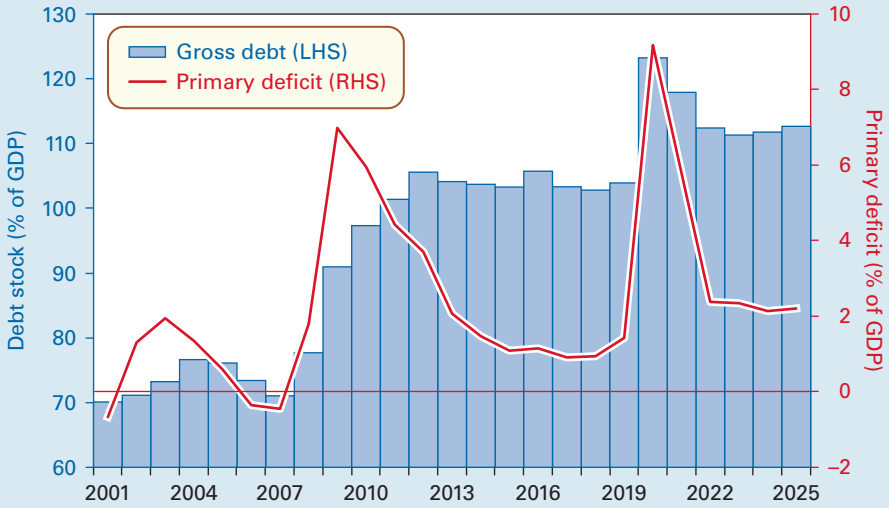
How can government expenditure and taxation be used to affect the level of economic activity?

In this section we begin by considering how expenditure and taxation flows automatically change across the business cycle and, in so doing, help to stabilise the economy. We then move on to discuss why governments might make deliberate changes to taxation and government expenditures and how effective they may be in managing the economy.

BOX 13.1 PRIMARY SURPLUSES AND SUSTAINABLE PUBLIC FINANCES

The fiscal arithmetic of government debt

General government primary deficit and debt in advanced economies



Notes: Forecasts from 2022; Group of advanced economies comprises 40 economies, *World Economic Outlook Database* (IMF, October 2022)

Source: Based on data from *Fiscal Monitor*, IMF eLibrary-Data (IMF, October 2022)

KI 23 p 173

The sustainability of a country's public finances is important in ensuring that governments can continue to support their economies through fiscal interventions when needed. For example, governments around the world undertook large fiscal interventions in response to the global financial crisis and the COVID-19 pandemic (see Box 13.2 for more on the magnitude of these interventions).

For the public finances to be sustainable, the public sector's debt-to-GDP ratio should not rise. To achieve this, it will need to run a particular primary balance-to-GDP ratio (*PB/Y*). The

Automatic fiscal stabilisers

To some extent, government expenditure and taxation will have the effect of *automatically* stabilising the economy. For example, as national income rises, the amount of tax people pay automatically rises. This rise in withdrawals from the circular flow of income helps to dampen down the rise in national income. This effect will be bigger if taxes are progressive (i.e. rise by a bigger percentage than national income). Some government transfers will have a similar effect. For example, the total paid in unemployment benefits will fall, if rises in national income cause a fall in unemployment. This again will have the effect of dampening the rise in national income.

Taxes whose revenues rise as national income rises and benefits that fall as national income rises are called **automatic stabilisers**.

Discretionary fiscal policy

Automatic stabilisers cannot *prevent* fluctuations; they merely reduce their magnitude. If there is a *fundamental* disequilibrium in the economy, or substantial fluctuations in national income, these automatic stabilisers will not be enough. The government may thus choose to *alter* the level of government expenditure or the rates of taxation. This is known as **discretionary fiscal policy**. For example, following the

Definition

Automatic fiscal stabilisers Tax revenues that rise and government expenditure that falls as national income rises. The more they change with income, the bigger the stabilising effect on national income.

EXPLORING ECONOMICS

primary balance is the difference between the public sector's revenues and its expenditures excluding the interest payments on its debts. A positive (negative) balance is consistent with a primary surplus (deficit). To understand the relationship between the primary balance and public-sector debt, consider the following debt sustainability rule:

$$\frac{PB}{Y} \geq \frac{D}{Y}(r - g)$$

The size of surplus (deficit) (*PB/Y*) will need to be larger (smaller), the higher the current debt-to-GDP ratio (*D/Y*), the higher the annual real rate of interest rate on public-sector debt (*r*) and the smaller the annual rate of real economic growth (*g*).

To illustrate the fiscal arithmetic, consider a country where the public-sector debt-to-GDP ratio (*D/Y*) is currently 0.5 (i.e. 50 per cent). Assume that the real interest rate on borrowing (*r*) is 3 per cent and the annual real rate of economic growth (*g*) is 1 per cent. To prevent the debt-to-GDP ratio (*D/Y*) from rising, the country would need to run a primary surplus of at least (≥) the equivalent to 1 per cent of GDP ($0.5 \times (3 - 1)$). If, however, the current debt-to-GDP ratio were currently 1 (i.e. 100 per cent), then it would need to run a larger primary surplus-to-GDP ratio of at least 2 per cent ($1 \times (3 - 1)$).

In the case where the economic growth rate (*g*) is greater than the real interest rate (*r*), the public sector can run a primary deficit without its debt-to-GDP ratio rising: in other words, if the right-hand side of (1) is negative, so too can be the left-hand side. This is important because many countries in the 2000s and 2010s experienced subdued inflationary pressures due to the impact of global trade and competition and the anchoring of inflationary expectations. As a result, central banks kept interest rates low, allowing governments to run large public-sector deficits.

Lower real interest rates make the fiscal arithmetic facing governments more favourable. However, as the chart shows, primary deficits and stocks of general government debt rose sharply in many countries as governments undertook significant discretionary fiscal interventions (see Box 13.2) to help mitigate the adverse economic effects arising from the global financial crisis and, later, the COVID-19 pandemic and then, in 2022, the rise in energy prices.

The primary deficit-to-GDP ratio across advanced economies rose to 7 per cent in 2009, while the debt-to-GDP ratio, which had been 71 per cent in 2007, had risen to 106 per cent by 2012. Declining primary deficits, a resumption in growth and low (or even negative) real interest rates then helped to stabilise the debt-to-GDP ratio for the remainder of the 2010s. However, the pandemic saw the primary deficit-to-GDP ratio rise to more than 9 per cent and the debt-to-GDP ratio to 123 per cent across advanced economies.

Some argue that the sustainability arithmetic is even more significant for countries in the eurozone. What might be the explanation for this argument?

From the AMECO database download data on the primary balance, as a percentage of GDP, for general government (net lending or borrowing excluding interest). Then, for a sample of up to five countries of your choice, plot a time series chart showing the primary balance to GDP ratio across time. Finally, compose a short briefing note summarising the patterns in your chart.

global financial crisis and the COVID-19 pandemic, many governments introduced measures designed to support the economy by stimulating aggregate demand. Box 13.2 discusses how we can measure the extent of such fiscal impulses and therefore changes in the fiscal stance of government.

TC 14 p 249

If government expenditure on goods and services (roads, health care, education, etc.) is raised, this will create a full multiplied rise in national income. The reason is that all the money gets spent and thus all of it goes to boosting aggregate demand.

Cutting taxes (or increasing benefits), however, will have a smaller effect on national income than raising government expenditure on goods and services by the same amount. The reason is that cutting taxes increases people's *disposable* incomes, of which only *part* will be spent. Part will be withdrawn into extra saving, imports and other taxes. In other words, not all the tax cuts will be passed on round the circular flow of income as extra expenditure. Thus if one-fifth of a cut in taxes is withdrawn and only four-fifths is spent, the tax multiplier will be only four-fifths as big as the government expenditure multiplier.

Definition

Discretionary fiscal policy Deliberate changes in tax rates or the level of government expenditure in order to influence the level of aggregate demand.

Pause for thought

Why will the multiplier effect of government transfer payments, such as child benefit, pensions and social security benefits, be less than the full multiplier effect from government expenditure on goods and services?

BOX 13.2 THE FISCAL IMPULSE

Assessing a country's fiscal stance

Discretionary policy and the fiscal impulse

If the government changes its expenditure or rates of taxation in order to affect aggregate demand, this is known as *discretionary fiscal policy*.

Discretionary fiscal measures played a significant role during both the global financial crisis and the COVID-19 pandemic as governments around the world attempted to support aggregate demand in the face of hugely negative pressures on private-sector spending. In the pandemic, for example, we saw countries adopt various job retention schemes. These allowed firms to furlough workers and apply for funds to cover some of their wages.

But how large were these fiscal stimuli and, more generally, how do we measure changes in the fiscal stance of governments?

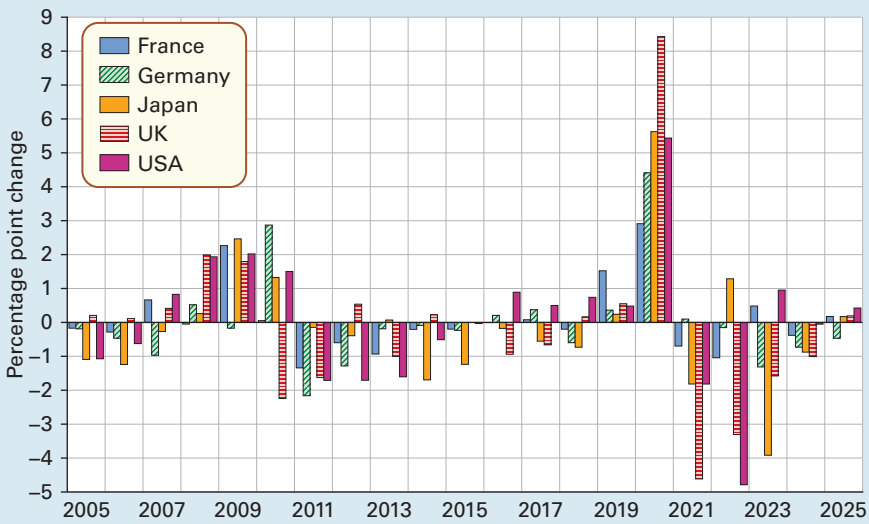
One way of measuring the magnitude of discretionary fiscal policy is through estimating what is known as the **fiscal impulse.** The concept should not be confused with that of fiscal multipliers. Multipliers, such as the government purchases (G) multiplier, measure the impact of fiscal changes on real national income. They are effectively measuring the *outcomes* arising from discretionary changes. In contrast, the fiscal impulse is concerned with the magnitude of the change in fiscal policy itself and thus with the size of the impulse whose effects are then transmitted through the economy.

Measuring the fiscal impulse

By measuring the fiscal policy impulse, we are able to analyse the extent to which a country's fiscal stance has tightened, loosened, or remained unchanged. To do so, we are attempting to capture changes in public spending and tax receipts that are not simply the automatic result of the economy moving from one point to another in the business cycle. Rather, the fiscal impulse shows *deliberate* policy changes which result in a structural change in public spending and/or taxation.

The fiscal impulse is therefore measured by changes in structural budget balances: i.e. cyclically-adjusted balances. A deterioration in a structural balance (a rise in the structural deficit or fall in the structural surplus) indicates a

Fiscal impulse



Notes: Data are the percentage point changes in the cyclically-adjusted general government primary deficit as a percentage of potential GDP; forecasts from 2022

Source: Based on data from Fiscal Monitor, IMF eLibrary-Data (IMF, October 2022)

The effectiveness of fiscal policy

There are two main problem areas with discretionary fiscal policy. The first concerns the *magnitude* of the effects of policy measures. If either government expenditure (G) or taxation (T) is changed, how much will *total* injections and withdrawals change? What will be the size of the resulting multiplier effect? How much will the change in aggregate demand affect output and employment, and how much will it affect prices?

The second concerns the *timing* of the effects. How quickly can policy be changed and how quickly will the changes affect the economy? KI 30 p 203

Problems of magnitude

Before changing government expenditure or taxation, the government will need to calculate the effect of any such change on national income, employment and inflation. Predicting these effects, however, is often very unreliable for a number of reasons.

EXPLORING ECONOMICS

loosening of the fiscal stance, whereas an improvement indicates a tightening.

A frequently-used measure of the fiscal impulse is the year-on-year change in the cyclically-adjusted general government primary deficit as a percentage of potential GDP. The general government primary deficit (see Box 13.1) captures the extent to which the receipts of general government fall short of its spending (excluding that on debt interest payments). By adjusting for the position of the economy in the business cycle, and thus any automatic changes in taxation or government expenditure, and excluding debt interest payments, the cyclically-adjusted primary deficit allows us to isolate more accurately the size of discretionary policy changes. A larger deficit or a smaller surplus indicates a fiscal loosening, while a smaller deficit or a larger surplus indicates a fiscal tightening.

The chart shows the fiscal impulse for a sample of countries from 2005. It is measured as the percentage point change in the cyclically-adjusted primary deficit to potential GDP ratio. Positive values indicate a fiscal loosening (a rise in the ratio) and negative values a fiscal tightening (a fall in the ratio).

The fiscal impulse in response to the pandemic

The chart clearly shows the extent of the *loosening* of countries' fiscal stance in 2020 in response to the COVID-19 pandemic. In the UK, for example, the cyclically-adjusted primary deficit to potential GDP ratio rose from 1.3 to 9.7 per cent. This represents a fiscal impulse of 8.4 per cent of GDP.

Although the primary deficit remained high across our sample of countries, a *tightening* of fiscal policy followed the waning of the pandemic, particularly so in the UK, USA and Japan. Yet the extent of the tightening was tempered by policy measures to limit the impact on households of the cost-of-living crisis and, in particular, rising energy and food prices.

The scale of the fiscal loosening in response to the pandemic is seen to dwarf that in response to the global financial crisis. The fiscal response to the financial crisis in the UK and the USA, which were both particularly badly affected, led to a cumulative increase in the cyclically-adjusted primary deficit-to-potential-GDP ratio in 2008 and 2009 of just 4 percentage points. However, as in the aftermath of the pandemic, concerns about the size of deficits led governments to tighten policy quite quickly. This meant that over the period from 2008 to 2014 the cumulative fiscal impulse in both countries was close to zero.

Above- and below-the-line measures of the fiscal impulse

Throughout the pandemic, the IMF maintained a database of measures by country. It classified measures as either *above* or *below* the line. The former involved additional spending or foregone tax revenues, while the latter includes loans, equity injections and loan guarantees.

As of 27 September 2021, the global discretionary fiscal policy response to the COVID-19 pandemic from January 2020 was estimated to have involved above- and below-the-line measures equivalent to 10.2 per cent and 6.2 per cent of global GDP in 2020, respectively and, therefore, fiscal support equivalent to 16.4 per cent of global GDP in 2020.

In the UK, above- and below-the-line measures were equivalent to 19.3 and 16.7 per cent of GDP in 2020, respectively, and therefore support was equivalent to a total of 36 per cent of GDP in 2020. For the USA, the figures were 25.5 per cent + 2.4 per cent = 27.9 per cent; for the EU as a whole they were 3.8 per cent + 6.7 per cent = 10.5 per cent, but for individual countries they were much higher, with Italy, the highest, having figures of 10.9 per cent + 35.3 per cent = 46.2 per cent; for Japan the figures were 16.7 per cent + 28.3 per cent = 45.1 per cent.

How is the concept of the fiscal impulse different from that of a fiscal multiplier?

Undertake desktop research into the fiscal policy interventions of a country of your choice that followed the pandemic.

Definition

Fiscal impulse A measure of the change in the fiscal stance arising from discretionary fiscal policy changes.

Predicting the effect of changes in government expenditure

A rise in government expenditure of £x may lead to a rise in total injections (relative to withdrawals) that is smaller than £x. This will occur if the rise in government expenditure *replaces* a certain amount of private expenditure. For example, a rise in expenditure on state education may dissuade some parents from sending their children to private schools. Similarly, an improvement in the National Health Service may lead to fewer people paying for private treatment.

Crowding out. Another reason for the total rise in injections being smaller than the rise in government expenditure is a phenomenon known as **crowding out**. If the government relies on **pure fiscal policy** – that is, if it does not finance an increase in the budget deficit by increasing the money supply – it will have to borrow the money from the non-bank private sector. It will

thus be competing with the private sector for finance and will have to offer higher interest rates. This will force the private sector too to offer higher interest rates, which may discourage firms from investing and individuals from buying on credit. Thus government borrowing crowds out private borrowing. In the extreme case, the fall in consumption and investment may completely offset the rise in government expenditure, with the result that aggregate demand does not rise at all.

Predicting the effect of changes in taxes

A cut in taxes, by increasing people's real disposable income, increases not only the amount they spend but also the amount they save. The problem is that it is not easy to predict the relative size of these two increases. In part it will depend on whether people feel that the cut in tax is only temporary, in which case they may simply save the extra disposable income, or permanent, in which case they may adjust their consumption upwards. More generally, it may depend on a broader set of variables, including confidence and financial well-being.

TC 11 p 61

Predicting the resulting multiplied effect on national income

Even if the government *could* predict the net *initial* effect on injections and withdrawals, the ultimate effect on national income will still be hard to predict for the following reasons:

TC 14 p 249

- The size of the *multiplier* may be difficult to predict, since it is difficult to predict how much of any rise in income will be withdrawn. In other words, it is difficult to predict the size of the *mpw*. For example, the amount of a rise in income that households save or consume will depend on their expectations about future price and income changes.

TC 11 p 61

- Induced investment through the *accelerator* (see pages 261–2) is also extremely difficult to predict. It may be that a relatively small fiscal stimulus will be all that is necessary to restore business confidence, and that induced investment will rise substantially. Similarly, rising confidence among financial institutions could see credit conditions relaxed with the *financial accelerator* (see pages 263–5) resulting in rising levels of investment. In such situations, fiscal policy can be seen as a 'pump primer'. It is used to start the process of recovery, and then the continuation of the recovery is left to the market. But for pump priming to work, people must believe that it will work. Confidence can change very rapidly and in ways that could not have been foreseen a few months earlier.
- Multiplier/accelerator interactions. If the initial multiplier and accelerator effects are difficult to estimate, their interaction will be virtually impossible to estimate. Small divergences in investment from what was initially predicted will become magnified as time progresses.

Random shocks

Forecasts cannot take into account the unpredictable, such as the attack on the World Trade Center in September 2001 or the COVID-19 pandemic in the 2020s. Even events that, with hindsight, should have been predicted, such as the banking crisis of 2007–9, often are not. Unfortunately, unpredictable or unpredicted events do occur and may seriously affect the government's fiscal policy.

Pause for thought

Give some other examples of 'random shocks' that could undermine the government's fiscal policy.

The problem of timing

Fiscal policy can involve considerable time lags. It may take time to recognise the nature of the problem before the government is willing to take action; tax or government expenditure changes take time to plan and implement – changes may have to wait until the next Budget to be announced and thus come into effect some time later; the effects of such changes take time to work their way through the economy via the multiplier and accelerator.

If time lags are long enough, fiscal policy could even be *de*stabilising. Expansionary policies taken to cure a recession may not come into effect until the economy has *already* recovered and is experiencing a boom. Under these circumstances, expansionary policies are quite inappropriate: they simply worsen the problems of overheating. Similarly, contractionary policies taken to prevent excessive expansion may not take effect until the economy has already peaked and is plunging into recession. The contractionary policies only deepen the recession.

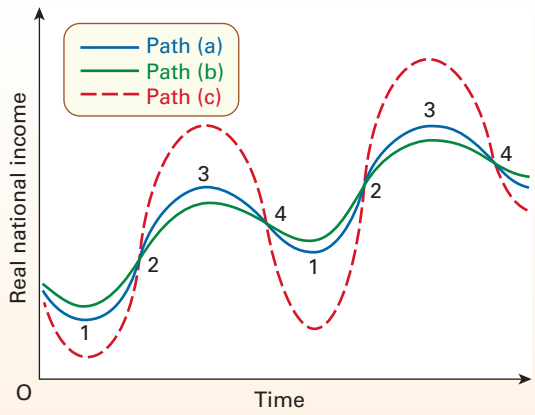
This problem is illustrated in Figure 13.3. Path (a) shows the course of the business cycle without government intervention. Ideally, with no time lags,

Definitions

Crowding out Where increased public expenditure diverts money or resources away from the private sector.

Pure fiscal policy Fiscal policy that does not involve any change in money supply.

Figure 13.3 Fiscal policy: stabilising or destabilising?



the economy should be dampened in stage 2 and stimulated in stage 4. This would make the resulting course of the business cycle more like path (b), or even, if the policy were perfectly stabilising, a line that purely reflected the growth in potential output. With time lags, however, contractionary policies taken in stage 2 may not come into effect until stage 4, and expansionary policies taken in stage 4 may not come into effect until stage 2. In this case the resulting course of the business cycle will be more like path (c). Quite obviously, in these circumstances 'stabilising' fiscal policy actually makes the economy *less* stable.

If the fluctuations in aggregate demand can be forecast, and if the lengths of the time lags are known, then all is not lost. At least the fiscal measures can be taken early and their delayed effects can be taken into account.

Fiscal rules

Given the problems of pursuing active fiscal policy, many governments in recent years took a much more passive approach. Instead of the policy being changed as the economy changes, countries applied a set of fiscal rules. These rules typically related to measures of government deficits and to the stock of accumulated debt. Taxes and government expenditure would then be planned to meet these rules.

However, designing rules for fiscal policy can be problematic. Economies are regularly hit by shocks. Some shocks may require more discretion than the current rules allow for. Hence, how flexible should any rules be?

Following the severe disruption to the global economy that occurred with the credit crunch of 2008, countries around the world resorted to discretionary fiscal policy to boost aggregate demand. Many temporarily abandoned fiscal rules. These were then generally reinstated around the world as the global economy pulled out of recession but were often more flexible, including, for example, so-called 'escape clauses'. However, the COVID-19 pandemic and later soaring energy prices again saw the adoption of extraordinary fiscal measures and the suspension of rules. Box 13.3 details the evolution of the Stability and Growth Pact: the EU's supranational fiscal framework.

In Section 13.4 we review the debate concerning constraints on a government's discretion over both its fiscal and monetary policies.

Recap

1. Automatic fiscal stabilisers are tax revenues that rise, and benefits that fall, as national income rises. They have the effect of reducing the size of the multiplier and thus reducing cyclical upswings and downswings.
2. Automatic stabilisers take effect as soon as aggregate demand fluctuates, but they can never remove fluctuations completely. They also create disincentives and act as a drag on recovery from recession.
3. Discretionary fiscal policy is where the government deliberately changes taxes or government expenditure in order to alter the level of aggregate demand. Changes in government expenditure on goods and services have a full multiplier effect. Changes in taxes and benefits, however, have a smaller multiplier effect as some of the tax/benefit changes will merely affect other withdrawals and thus have a smaller net effect on the consumption of domestic goods.
4. There are problems in predicting the magnitude of the effects of discretionary fiscal policy. Expansionary fiscal policy can act as a pump primer and stimulate increased private expenditure, or it can crowd out private expenditure. The extent to which it acts as a pump primer depends crucially on business confidence – something that is very difficult to predict beyond a few weeks or months. The extent of crowding out depends on monetary conditions and monetary policy.
5. There are various time lags involved with fiscal policy, which make it difficult to use fiscal policy to 'fine-tune' the economy.
6. In recent years, many governments around the world preferred a more passive approach towards fiscal policy. Targets were set for one or more measures of the public-sector finances, and then taxes and government expenditure were adjusted so as to keep to the target.
7. Nevertheless, in extreme circumstances, as occurred in 2008/9, and during the coronavirus pandemic, governments were prepared to abandon rules and give a fiscal stimulus to their economies.

BOX 13.3 THE EVOLUTION OF THE STABILITY AND GROWTH PACT IN THE EU

A supranational fiscal framework

Preparing for the euro

In signing the Maastricht Treaty in 1992, the EU countries agreed that to be eligible to join the single currency (i.e. the euro), they should have sustainable deficits and debts. This was interpreted as follows: the general government deficit should be no more than 3 per cent of GDP and general government debt should be no more than 60 per cent of GDP, or should at least be falling towards that level at a satisfactory pace.

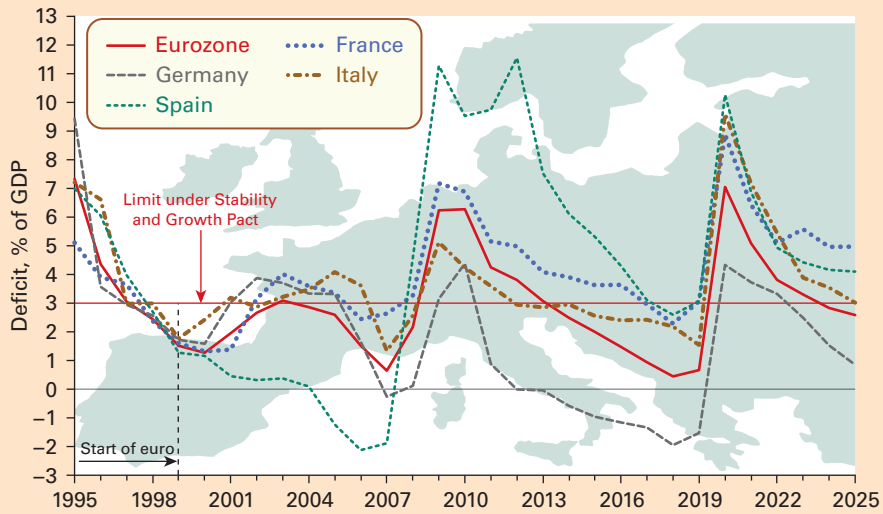
But in the mid-1990s, several of the countries that were subsequently to join the euro had deficits and debts substantially above these levels (see chart). Getting them down proved a painful business. Government expenditure had to be cut and taxes increased. These fiscal measures, unfortunately, proved to be powerful! Unemployment rose and growth remained low.

The EU Stability and Growth Pact (SGP)

In June 1997, at the European Council meeting in Amsterdam, the EU countries agreed a Stability and Growth Pact (SGP). This stated that member states should seek to balance their budgets (or even aim for a surplus) averaged over the course of the business cycle, and that deficits should not exceed 3 per cent of GDP in any one year. A country's deficit was permitted to exceed 3 per cent only if its GDP had declined by at least 2 per cent (or 0.75 per cent with special permission from the Council of Ministers). Otherwise, countries with deficits exceeding 3 per cent were required to make deposits of money with the European Central Bank. Under the Pact's Excessive Deficit Procedure, these would then become fines if the excessive budget deficit were not eliminated within two years.

There were two main aims of targeting a zero-budget deficit over the business cycle. The first was to allow automatic

General government deficits in the eurozone



Note: Data from 2022 based on forecast

Source: Based on data in *World Economic Outlook Database* (IMF, October 2022)

13.3 MONETARY POLICY

How can the supply of money and interest rates be controlled?

The Bank of England's Monetary Policy Committee meets eight times per year to set Bank Rate. The event gets considerable media coverage especially when a change is expected.

Changes in interest rates have gained a central significance in macroeconomic policy. And it is not just in the UK. Whether it is the European Central Bank setting interest rates for the eurozone countries, or the Federal Reserve Bank setting US interest rates, or any other central bank around the world choosing what the level of interest rates should be, monetary policy is seen as having a major influence on a whole range of macroeconomic indicators.

But is monetary policy simply the setting of interest rates? In reality, it involves the central bank intervening in the money market to ensure that the

CASE STUDIES & APPLICATIONS

stabilisers to work without 'bumping into' the 3 per cent deficit ceiling in years when economies were slowing. The second was to allow a reduction in government debts as a proportion of GDP (assuming that GDP grew on average at around 2–3 per cent per year).

From 2002, with slowing growth, Germany, France and Italy breached the 3 per cent ceiling (see chart). By 2007, however, after two years of relatively strong growth, deficits had been reduced well below the ceiling.

But then the credit crunch hit. As the EU economies slowed, so deficits rose. To combat the recession, in November 2008 the European Commission announced a €200 billion fiscal stimulus plan, mainly in the form of increased public expenditure. Of this sum, €170 billion would come from member governments and €30 billion from the EU, amounting to a total of 1.2 per cent of EU GDP. The money would be for a range of projects, such as job training, help to small businesses, developing green energy technologies and energy efficiency. Most member governments quickly followed by announcing how their specific plans would accord with the overall plan.

The combination of the recession and the fiscal measures pushed most EU countries' budget deficits well above the 3 per cent ceiling (see chart). The recession in EU countries deepened markedly in 2009, with GDP declining by 4.5 per cent in the eurozone as a whole, and by 5.6 per cent in Germany, 5.5 per cent in Italy, 3.6 per cent in Spain and 2.9 per cent in France. Consequently, the deficits were not seen to breach SGP rules.

In some cases, countries' public finances deteriorated unsustainably. Following high-profile rescue packages to Greece, Ireland and Portugal involving the International Monetary Fund and EU, the EU established the European Stability Mechanism (ESM) for eurozone countries in financial difficulties. The ESM became operational in October 2012 and can provide loans or purchase the countries' bonds in the primary market.

The Fiscal Compact

With many countries experiencing burgeoning deficits and some countries requiring financial assistance, the SGP was no longer seen as a credible vehicle for constraining deficits: it needed reform. The result was an intense period of negotiation that culminated in early 2012 with a new intergovernmental treaty on limiting spending and borrowing.

The treaty, known as the Fiscal Compact, requires that national governments not only abide by the excessive deficit procedure of the SGP but also keep structural deficits (i.e. deficits that would remain even if economies were operating at potential output) no higher than 0.5 per cent of GDP.

In the cases of countries with a debt-to-GDP ratio significantly below 60 per cent, the structural deficit is permitted to reach 1 per cent of GDP. Where the debt-to-GDP ratio exceeds 60 per cent, countries should reduce this excess at an average rate of one-twentieth per year. A failure to do so can lead to a fine of up to 0.1 per cent of GDP payable to the ESM.

The COVID-19 pandemic and the 'escape clause'

The average structural deficit in the eurozone fell from 5.0 per cent of GDP in 2010 to 0.7 per cent in 2019. Nonetheless, many countries still had structural deficits in excess of the permitted levels.

However, in May 2020 the European Commission announced that it was activating the 'escape clause' introduced as part of the reforms of the SGP to enable member states to respond to the pandemic. The cyclically-adjusted primary balance to potential GDP ratio in the eurozone fell from +0.7 per cent in 2019 to –4.7 per cent in 2020, representing a fiscal impulse (see Box 13.2) of 5.4 per cent of GDP. (This compares with just 1.4 per cent in 2009.) The escape clause was expected continue through to at least 2023.

What effects will an increase in government investment expenditure have on public-sector debt (a) in the short run; (b) in the long run?

From the IMF World Economic Outlook Database download data on the actual and structural budget balances, as a percentage of GDP, for general government (net lending). Then, for Germany and the UK, plot a time-series chart showing both balances across time. Finally, compose a short briefing note summarising the patterns in your chart.

interest rate that has been announced is also the *equilibrium* interest rate.

The policy setting

TC 7 p 26

In framing its monetary policy, the government must decide on what the goals of the policy are. Is the aim simply to control inflation, or does the government wish also to affect output and employment, or does it want to control the exchange rate?

The government also has to decide the role of the central bank in carrying out monetary policy. There are three possible approaches.

In the first, the government both sets the policy and decides the measures necessary to achieve it. Here the government would set the interest rate, with the central bank simply influencing money markets to achieve this rate. This first approach was used in the UK before 1997.

The second approach is for the government to set the policy *targets*, but for the central bank to be given independence in deciding interest rates. This is the approach adopted in the UK today. The government has set a target rate for CPI inflation of 2 per cent, but then the MPC is free to choose the rate of interest.

The third approach is for the central bank to be given independence not only in carrying out policy,

but in setting the policy targets itself. The ECB, within the statutory objective of maintaining price stability in the euro area, has decided on the target of keeping inflation at 2 per cent over the medium term.

Finally, there is the question of whether the government or central bank should take a long-term or short-term perspective. Should it adopt a target for inflation or money supply growth and stick to it come what may? Or should it adjust its policy as circumstances change and attempt to 'fine-tune' the economy?

We shall be looking primarily at *short-term* monetary policy: that is, policy used to keep to a set target for inflation or money supply growth, or policy used to smooth out fluctuations in the business cycle.

It is important first, however, to take a longer-term perspective. Governments generally want to prevent an excessive growth in the money supply over the longer term. Likewise, they want to ensure that money supply grows enough and that there is not a shortage of credit, such as that during the credit crunch. If money supply grows too rapidly, then inflation is likely to become embedded in the economy, which will feed inflationary expectations; if money supply grows too slowly, or even falls, then recession is likely to result, investment may fall and long-term growth may be subdued.

Control of the money supply over the medium and long term

In Section 11.3 we identified two major sources of monetary growth: (a) banks choosing to hold a lower liquidity ratio (probably in response to an increase in the demand for loans); (b) public-sector borrowing financed by borrowing from the banking sector. If the government wishes to restrict monetary growth over the longer term, it could attempt to control either or both of these.

Banks' liquidity ratio

The central bank could impose a statutory **minimum reserve ratio** on the banks, *above* the level that banks would otherwise choose to hold. Such ratios come in various forms. The simplest is where the banks are required to hold a given minimum percentage of deposits in the form of cash or deposits with the central bank.

The effect of a minimum reserve ratio is to prevent banks choosing to reduce their cash or liquidity ratio and creating more credit. This was a popular approach of governments in many countries in the past. Some countries imposed very high ratios indeed in their attempt to slow down the growth in the money supply.

A major problem with imposing restrictions of this kind is that banks may find ways of getting around them. After all, banks would like to lend and customers would like to borrow. It is very difficult to regulate and police every single part of countries' complex financial systems.

Nevertheless, attitudes changed substantially after the excessive lending of the mid 2000s. The expansion of credit had been based on 'liquidity' achieved through secondary marketing between financial institutions and the growth of securitised assets containing sub-prime debt (see Box 11.3). After the credit crunch and the need for central banks or governments to rescue ailing banks, such as Northern Rock and later the Royal Bank of Scotland in the UK and many other banks around the world, there were calls for greater regulation of banks to ensure that they had sufficient capital and operated with sufficient liquidity and that they were not exposed to excessive risk of default. As we saw in Section 11.2, several measures were taken.

KI 15 p 80

Public-sector deficits

Section 11.3 showed how government borrowing tends to lead to an increase in money supply. To prevent this, public-sector deficits must be financed by selling *bonds* (as opposed to bills, which could well be taken up by the banking sector, thereby increasing money supply). However, to sell extra bonds the government will have to offer higher interest rates. This will have a knock-on effect on private-sector interest rates. The government borrowing will thus crowd out private-sector borrowing and investment. This is known as **financial crowding out**.

If governments wish to reduce monetary growth and yet avoid financial crowding out, they must therefore reduce the size of public-sector deficits.

The less successful a government is in controlling the public-sector deficit, the more it will have to borrow through bond issue, to prevent money supply growing too fast. This will mean high interest rates and the problem of crowding out, and a growing burden of public-sector debt with interest on it that has to be paid from taxation, from further cuts in government expenditure, or from further borrowing.

It is partly for this reason that many countries have attempted to constrain fiscal policy choices by applying fiscal rules or frameworks, such as the Stability and Growth Pact in the eurozone (see Box 13.3).

Pause for thought

How could long-term monetary growth come about if the government persistently ran a public-sector surplus?

Short-term monetary measures

Assume that inflation is above its target rate and that the central bank wishes to operate a tighter monetary policy in order to reduce aggregate demand and so the rate of inflation. What can it do?

When analysing expenditure decisions, it is typically *real* interest rates (r) that are important. Consumers and producers are interested in the additional future volumes of consumption and production respectively that their investment, borrowing or saving today will enable them to enjoy.

TC 5 p 17

The *realised* ('*ex post*') real rate of interest (r) received on savings or paid on borrowing is the nominal (actual) interest rate (i) *less* the rate of inflation (π). However, for savers and borrowers it is the future rate of inflation that is relevant in their decision making. Of course, future inflation rates cannot be known with certainty and, instead, people must form *expectations* of inflation. Hence, the real interest rate when the decision is made ('*ex ante*') is the nominal interest rate (i) *less* the *expected* rate of inflation (π^e).

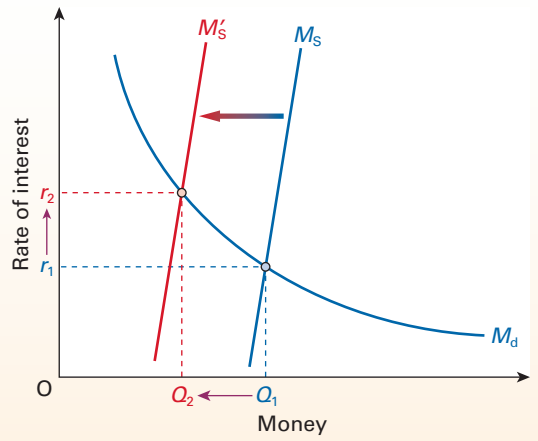
TC 5 p 17

We will assume for ease of argument that, in the short term, the central bank is able to take the expected rate of inflation (π^e) as given. Consequently, any change in the nominal rate of interest (i) will be matched by an equivalent change in the real rate of interest (r).

For any given supply of money (M_s) there will be a particular equilibrium real rate of interest at any one time: where the supply of money (M_s) equals the demand for money (M_d). This is shown as r_1 in Figure 13.4.

Thus to operate a tighter monetary policy, the authorities can do one of the following:

Figure 13.4 The demand for and supply of money



- Reduce money supply and accept whatever equilibrium interest rate results. Thus if money supply is reduced to Q_2 in Figure 13.4, a new higher rate of interest, r_2, will result.
- Raise interest rates to r_2 and then manipulate the money supply to reduce it to Q_2. The more endogenous the money supply is, the more this will occur automatically through banks adjusting credit to match the lower demand at the higher rate of interest and the less the central bank will have to take deliberate action to reduce liquidity.
- Keep interest rates low (at r_1), but also reduce money supply to a level of Q_2. The trouble here is that the authorities cannot both control the money supply *and* keep interest rates down without running into the problem of disequilibrium. Since the demand for money now exceeds the supply by $Q_1 - Q_2$, some form of credit rationing would have to be applied.

Credit rationing was widely used in the past, especially during the 1960s. The aim was to keep interest

Definitions

Minimum reserve ratio A minimum ratio of cash (or other specified liquid assets) to deposits (either total or selected) that the central bank requires banks to hold.

Financial crowding out Where an increase in government borrowing diverts money away from the private sector.

rates low, so as not to discourage investment, but to restrict credit to more risky business customers and/or to consumers. In the UK, the Bank of England could order banks to abide by such a policy, although in practice it always relied on persuasion. The government also, from time to time, imposed restrictions on hire-purchase credit, by specifying minimum deposits or maximum repayment periods.

Such policies were progressively abandoned around the world from the early 1980s. They were seen as stifling competition and preventing efficient banks from expanding. Hire-purchase controls may badly hit certain industries (e.g. cars and other consumer durables), whose products are bought largely on hire-purchase credit. What is more, with the deregulation and globalisation of financial markets up to 2007, it had become very difficult to ration credit. If one financial institution were controlled, borrowers could simply go elsewhere.

With the excessive lending in sub-prime markets that had triggered the credit crunch of 2007–9, however, there were calls around the world for tighter controls over bank lending. But this was different from credit rationing as we have defined it. In other words, tighter controls, such as applying counter-cyclical buffers of capital to all banks, would be used to prevent reckless behaviour by banks, rather than to achieve a particular level of money at a lower rate of interest.

We thus focus on controlling the money supply and controlling interest rates.

Techniques to control the money supply

There are four possible techniques that a central bank could use to control money supply. They have one major feature in common: they involve manipulating the liquid assets of the banking system. The aim is to influence the total money supply by affecting the amount of credit that banks can create.

Open-market operations. **Open-market operations** are the most widely used of the four techniques around the world. They alter the monetary base (cash and reserves in the central bank). This then affects the amount of credit banks can create and hence the level of broad money (M4 in the UK; M3 in the eurozone).

Open-market operations have historically involved the sale or purchase by the central bank of government securities (bonds or bills) in the open market. These sales or purchases are *not* in response to changes in the public-sector deficit. Rather, they are being conducted to implement monetary policy. Hence, they are best understood in the context of an unchanged deficit.

Alternatively, open-market operations can involve either 'repurchase agreements (repos)' or 'reverse repurchase agreements'. Repos allow the central bank to *increase* the money supply (temporarily) by buying government securities (or other eligible securities) from financial institutions, thereby supplying them with reserves. The financial institution agrees to buy the securities back at a later date. If, however, the central bank wanted to *reduce* the money supply on a temporary basis, it will use reverse repos, which involve selling government securities (thereby withdrawing reserves), with an agreement to buy them back later. Details of how open-market operations work in the UK are given in Box 13.4.

TC 14 p 249

Pause for thought

Explain how open-market operations could be used to increase the money supply.

Central bank lending to the banks. The central bank in most countries is prepared to provide extra money to banks (through gilt repos, rediscounting bills or straight loans). In some countries, it is the policy of the central bank to keep its interest rate to banks below market rates, thereby encouraging banks to borrow (or sell back securities) whenever such facilities are available. By cutting back the amount it is willing to provide, the central bank can reduce banks' liquid assets and hence the amount of credit they can create.

In other countries, such as the UK and the eurozone countries, it is normally not so much the amount of money made available that is controlled, but rather the rate of interest (or discount). The higher this rate is relative to other market rates, the less banks will be willing to borrow, and the lower, therefore, the monetary base will be. Raising this rate, therefore, has the effect of reducing the money supply.

In response to the credit crunch of the late 2000s, central banks in several countries extended their willingness to lend to banks. The pressure on central banks to act as the 'liquidity backstop' grew as the inter-bank market ceased to function effectively in distributing reserves and, hence, liquidity between financial institutions. As a result, inter-bank rates rose sharply relative to the policy rate. Increasingly, the focus of central banks was on providing the necessary liquidity to ensure the stability of the financial system.

Definition

Open-market operations The sale (or purchase) by the authorities of government securities in the open market in order to reduce (or increase) money supply.

Yet, at the same time, by providing more liquidity, central banks were ensuring monetary policy was not being compromised. The additional liquidity was needed to alleviate the upward pressure on market interest rates.

Funding. Rather than focusing on controlling the monetary base (as in the case of the above two techniques), an alternative is for the authorities (the Debt Management Office in the UK) to alter the overall liquidity position of the banks. An example of this approach is a change in the balance of **funding** government debt. To reduce money supply, the authorities issue more bonds and fewer bills. Banks' balances with the central bank will be little affected, but to the extent that banks hold fewer bills, there will be a reduction in their liquidity and hence a reduction in the amount of credit created. Funding is thus the conversion of one type of government debt (liquid) into another (illiquid).

One problem with this approach is that bonds are likely to command a higher interest rate than bills. By switching from bills to bonds, the government will be committing itself to these interest rates for the life of the bond.

Variable minimum reserve ratios. In some countries (such as the USA), banks are required to hold a certain proportion of their assets in liquid form. The assets that count as liquid are known as 'reserve assets'. These include assets such as balances in the central bank, bills of exchange, certificates of deposit and money market loans. The ratio of such assets to total liabilities is known as the *minimum reserve ratio*. If the central bank raises this ratio (in other words, requires the banks to hold a higher proportion of liquid assets), then banks will have to reduce the amount of credit they grant. The money supply will fall.

Difficulties in controlling money supply

Targets for the growth in broad money were an important part of UK monetary policy from 1976 to 1985. Money targets were then abandoned in the UK and have not been used since. In the eurozone, until 2003 the European Central Bank adopted a reference value for the growth of M3. This, however, was only a guideline and not a strict target. If, however, a central bank did choose to target money supply as its main monetary policy, how would the policy work?

Assume that money supply is above target and that the central bank wishes to reduce it. It would probably use open-market operations: i.e. it would sell more bonds or bills. The purchasers of the bonds or bills would draw liquidity from the banks. Banks would then supposedly be forced to cut down on the credit they create. But is it as simple as this?

The problem is that banks will normally be unwilling to cut down on loans if people want to borrow – after all, borrowing by customers earns profits for the banks. Banks can always 'top up' their liquidity by borrowing from the central bank and then carry on lending. True, they will have to pay the interest rate charged by the central bank, but they can pass on any rise in the rate to their customers.

The point is that as long as people *want* to borrow, banks and other financial institutions will normally try to find ways of meeting the demand. In other words, in the short run at least, the supply of money is to a large extent demand determined. It is for this reason that central banks prefer to control the *demand* for money by controlling interest rates (backed up, normally, by open-market operations).

As we shall see in Box 13.6, there are similar difficulties in *expanding* broad money supply by a desired amount. Following the credit crunch and later the pandemic, various central banks around the world engaged in a process of quantitative easing. The process results in an increase in the monetary base (narrow money): banks' liquidity increases. But just how much this results in an increase in broad money depends on the willingness of banks to lend and customers to borrow. In the recessionary climate after 2008, confidence was low. Much of the extra liquidity remained in banks and the money multiplier fell (see Figure 11.4 on page 324). The growth of M4 in the UK fell sharply during 2009, despite quantitative easing, and remained weak throughout the first half of the 2010s (see Figure 11.5 on page 325).

Techniques to control interest rates

The approach to monetary control today in most countries is to focus directly on interest rates. Normally an interest rate change will be announced, and then open-market operations will be conducted by the central bank to ensure that the money supply is adjusted so as to make the announced interest rate the *equilibrium* one.

Central banks look to affect *real* interest rates through their monetary operations to affect nominal interest rates. Assuming inflationary expectations are constant in the short term, then, by announcing a particular change in nominal rates, this results in an equivalent change in real interest rates. Thus, in Figure 13.4 (on page 389), the central bank might announce a rise in nominal interest rates which corresponds to an equivalent rise in real interest rates from r_1 to r_2 and then conducts open-market operations to ensure that the money supply is reduced from Q_1 to Q_2.

TC 5 p 17

Definition

Funding Where the authorities alter the balance of bills and bonds for any given level of government borrowing.

BOX 13.4 THE OPERATION OF MONETARY POLICY IN THE UK — CASE STUDIES & APPLICATIONS

Managing the reserves

The Bank of England (the 'Bank') does not normally attempt to control money supply directly. Instead it seeks to control interest rates. The Monetary Policy Committee (MPC) of the Bank of England usually meets eight times per year to decide on Bank Rate. Changes in the Bank Rate are then intended to affect the whole structure of interest rates in the economy, from inter-bank rates to bank deposit rates and rates on mortgages and business loans. The main way in which it does this is by paying interest at the Bank Rate on deposits placed overnight with the Bank: i.e. on the reserve accounts of financial institutions.

Asset purchases

Ordinarily, the Bank of England uses *regular* open-market operations (OMOs) to keep short-term interest rates close to the Bank Rate by changing the supply of money so as to match the demand at Bank Rate. This helps to ensure that changes in the demand for reserves of financial institutions do not result in significant volatility in short-term market rates and in deviations from Bank Rate.

However, from March 2009 the Bank of England began deliberately injecting narrow money in a process known as *quantitative easing* to meet its inflation-rate target. This involved large-scale asset purchases, mainly government bonds, from financial institutions (see Box 13.6 on page 396). The purchases were made with newly created (electronic) money, which was credited to banks as reserve assets.

Between March 2009 and July 2012 the Bank of England injected £375 billion through the purchase of gilts. Then, following the EU referendum, the Bank purchased £60 billion of additional government bonds and £10 billion of corporate bonds between August 2016 and April 2017. Then, between March 2020 and December 2021, in response to the COVID-19 pandemic, the Bank increased its holdings of government and corporate bonds by a further £450 billion, taking its total asset purchases to £895 billion.

In December 2021 the Bank stopped buying new bonds. Then in February 2022, in response to significant inflationary pressures, the MPC decided that it would begin reducing its stock of assets in the autumn. This marked a period of *quantitative tightening*. First, it ceased to reinvest the monies from maturing bonds and then began a programme of bond sales. In effect, this meant the return to the Bank of England of money previously injected into the economy through asset purchases and its removal from the financial system.

Floors and ceilings

During the period of quantitative easing, the aggregate level of banks' reserves became determined by the quantity of asset purchases. Given the abundance of reserves that were therefore available, and which could be deposited with the Bank and earn Bank Rate, wholesale interest rates settled close to Bank Rate. The ability of banks to borrow money in the market and then deposit it overnight and earn Bank Rate prevented market rates falling significantly below Bank Rate. The abundance of reserves thus created a *floor* in short-term interest rates.

Downward pressure on interest rates continued to be exerted through the Bank's 'operational standing facilities'. These allow individual banks to borrow overnight at 25 basis points (0.25 percentage points) above Bank Rate, thereby creating a *ceiling* to short-term interest rates by providing an alternative source of borrowing for financial institutions. However, during periods of quantitative easing, the abundance of reserves meant that the monetary policy framework was consistent with a *floor system*.

Term Funding

In March 2020 in response to the COVID-19 pandemic, the Bank of England cut Bank Rate from 0.75 per cent to 0.1 per cent. However, with such low interest rates it feared that banks and building societies might find it difficult to reduce their rates on savings accounts much further, thus making it difficult to reduce lending rates. In other words, the pass-through effect from policy rates to market interest rates may cease to operate effectively.

To help ensure that the cut in Bank Rate was passed on to borrowers, financial institutions could borrow from the Bank of England to grant four-year loans at rates 'at or very close to Bank Rate'. The quantity of funds made available to banks under this 'Term Funding Scheme' was at least 10 per cent of banks' total real economy lending. This percentage could be higher for banks that increased lending, especially to small and medium-sized enterprises (SMEs). The scheme was known as 'TFSME' (term funding for small and medium-sized enterprises).

The TFSME mirrored a Term Funding Scheme introduced in August 2016 following the outcome of the EU-referendum vote. That scheme, which closed to new lending in February 2018, supported £127 billion of loans. The drawdown period for TFSME loans ended in October 2021. By May 2022 the scheme had supported loans to the value £192 billion.

Quantitative tightening

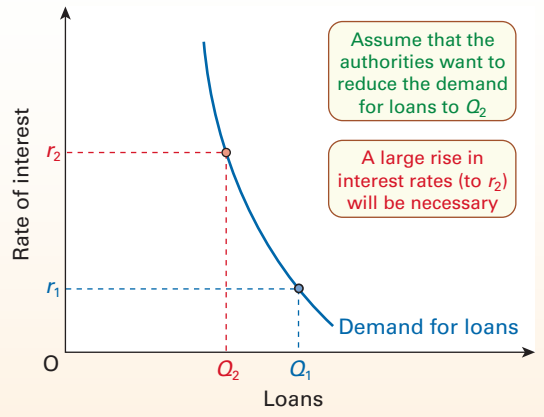
With the inflationary spike of 2021–22, the period of quantitative easing that had started after the global financial crisis came to an end. This heralded in a period of quantitative tightening, with the Bank reducing money supply as necessary to support rises in Bank Rate. At some point, banks' reserves would fall below the levels demanded. The Bank of England would then supply (lend them) reserves through regular OMOs, as had been the case before the financial crisis. Case 13.10 on the student website details the monetary policy framework that operated previously.

The process of quantitative tightening was delayed, however, in response to a large package of energy subsidies and tax cuts announced under the leadership of Liz Truss in September 2022. This would have involved a huge increase in government borrowing. Existing bond prices fell rapidly and, in order to stabilise the market, the Bank was forced into bond purchases. Many of the policies were reversed when the government of Rishi Sunak laid out a programme of fiscal consolidation. Quantitative tightening began in November 2022.

Assume that the Bank of England wants to raise interest rates. Trace through the process by which it achieves this.

Using the Bank of England Statistical Interactive Database, download data for the Bank Rate paid on reserves (series YWMB47D). Create a chart of the series in Excel and then write a short summary explaining both what Bank Rate is and what your chart shows.

Figure 13.5 An inelastic demand for loans



Let us assume that the central bank decides to raise interest rates. What does it do? In general, it will seek to keep banks short of liquidity. This will happen automatically on any day when tax payments by banks' customers exceed the money they receive from government expenditure. This excess is effectively withdrawn from banks and ends up in the government's account at the central bank. Even when this does not occur, issues of government debt will effectively keep the banking system short of liquidity, at least in the short term.

This 'shortage' can then be used as a way of forcing through interest rate changes. Banks will obtain the necessary liquidity from the central bank through repos or by selling it bills. The same logic applies if the central bank wants to lower interest rates: it creates a surplus through reverse repo operations or by buying bills. Following the financial crisis, and again during the COVID-19 pandemic, many central banks adopted an aggressive form of OMOs, known as 'quantitative easing'. This involved large-scale asset purchases (see Boxes 13.4 and 13.6).

The effectiveness of changes in interest rates

Even though central bank adjustment of the repo rate is the current preferred method of monetary control in most countries, it is not without its difficulties. The problems centre on the nature of the demand for loans. If this demand is (a) unresponsive to interest rate changes or (b) unstable because it is significantly affected by other determinants (such as anticipated income or foreign interest rates), then it will be very difficult to control by controlling the rate of interest.

TC 10 p 49

Problem of an inelastic demand for loans. If the demand for loans is inelastic, as in Figure 13.5, any attempt to reduce demand (e.g. from Q_1 to Q_2) will involve large rises in interest rates (r_1 to r_2). The problem will be compounded if the demand shifts to the right, due, say, to a consumer-spending boom. High interest rates lead to the following problems:

- They may discourage investment (as opposed to current consumption) and hence long-term growth.
- They add to the costs of production, to the costs of house purchase and generally to the cost of living. They are thus cost inflationary.
- They are politically unpopular, since the general public does not like paying higher interest rates on overdrafts, credit cards and mortgages.
- The authorities may need to ensure a sufficient supply of longer-term securities so that liquidity can be constrained. This could commit the government to paying high rates on these bonds for some time.
- High interest rates encourage inflows of money from abroad. This drives up the exchange rate. (We examine this in Chapter 15.) A higher exchange rate makes domestic goods expensive relative to goods made abroad. This can be very damaging for export industries and industries competing with imports. Many firms in the UK suffered badly between 1997 and 2007 from a high exchange rate, caused partly by higher interest rates in the UK than in the eurozone and the USA.

Evidence suggests that the demand for loans may indeed be quite inelastic, especially in the short run. Although investment plans may be curtailed by high interest rates, borrowing to finance current expenditure by many firms cannot easily be curtailed. Similarly, while householders may be discouraged from taking on new mortgages, they may find it difficult to reduce current expenditure as a way of reducing their credit-card debt. What is more, although high interest rates may discourage many firms from taking out long-term fixed-interest loans, some firms may merely switch to shorter-term variable-interest loans.

TC 11 p 61

Problem of an unstable demand. Accurate monetary control requires the authorities to be able to predict the demand curve for money (in Figure 13.5). Only then can they set the appropriate level of interest rates. Unfortunately, the demand curve may shift unpredictably, making control very difficult. The major reason is *speculation*:

- If people think interest rates will rise and bond prices fall, they will in the meantime demand to hold their assets in liquid form. The demand for money will rise.
- If people think exchange rates will rise, they will hold the domestic currency while it is still relatively cheap. The demand for money will rise.
- If people think the rate of inflation will rise, the transactions demand for money may rise. People plan to spend more while prices are still relatively low.

BOX 13.5 MONETARY POLICY IN THE EUROZONE

The role of the ECB

The European Central Bank (ECB) is based in Frankfurt and is charged with operating the monetary policy of those EU countries that have adopted the euro. Although the ECB has the overall responsibility for the eurozone's monetary policy, the central banks of the individual countries, such as the Bank of France and Germany's Bundesbank, were not abolished. They are responsible for distributing euros and for carrying out the ECB's policy with respect to institutions in their own countries. The whole system of the ECB and the national central banks is known as the European System of Central Banks (ESCB).

In operating the monetary policy of a 'euro economy' roughly the size of the USA, and in being independent from national governments, the ECB's power is enormous and is equivalent to that of the Fed. So, what is the structure of this giant on the European stage, and how does it operate?

The structure of the ECB

The ECB has two major decision-making bodies: the Governing Council and the Executive Board.[1]

The Governing Council consists of the members of the Executive Board and the governors of the central banks of each of the eurozone countries. The Council's role is to set the main targets of monetary policy and to take an oversight of the success (or otherwise) of that policy. It also sets interest rates at six-weekly meetings. Decisions are by simple majority. In the event of a tie, the president has the casting vote.

The Executive Board consists of a president, a vice-president and four other members. Each serves for an eight-year, non-renewable term. The Executive Board is responsible for implementing the decisions of the Governing Council and for preparing policies for the Council's consideration. Each member of the Executive Board has a responsibility for some particular aspect of monetary policy.

ECB independence

The ECB is one of the most independent central banks in the world. It has very little formal accountability to elected politicians. Although its President can be called before the European Parliament, the Parliament has virtually no powers to influence the ECB's actions.

Until its January 2015 meeting, its deliberations were secret and no minutes of Council meetings were published. Subsequently, an account of meetings is published (usually with a lag of around 2 weeks) with an explanation of the policy stance. However, the minutes do not include details of how Council members voted or of future policy intentions, unlike the minutes published by the Bank of England which, from 2015, are available at the time of the policy announcement.

Inflation rate target

The overall responsibility of the ECB is to achieve price stability in the eurozone. Since July 2021, the ECB has adopted a symmetric medium-term target of 2 per cent. This replaced the former target of below but close to 2 per cent over the medium term. It is a weighted average rate for all the members of the eurozone, not a rate that has to be met by every member individually.

The ECB attempts to 'steer' short-term interest rates to meet its inflation rate target. From November 2022, the rates were as follows: 2.00 per cent for the main 'refinancing operations' of the ESCB (i.e. the minimum rate of interest at which liquidity is offered once per week to 'monetary financial institutions' (MFIs) by the ESCB); a 'marginal lending' rate of 2.25 per cent (for providing overnight support to the MFIs); and a 'deposit rate' of 1.50 per cent (the rate paid to MFIs for depositing overnight surplus liquidity with the ESCB).

The ECB began operating a *negative* deposit rate from June 2014 and this persisted until July 2022. A negative rate meant that banks were being charged for holding money with the ECB rather than lending it. The intention was to encourage banks to lend to each other or to households and businesses and, consequently, stimulate the economy.

The operation of monetary policy

The ECB sets a minimum reserve ratio for eurozone banks. The ratio is designed primarily to prevent excessive lending and hence the need for excessive borrowing from the central bank or from other financial institutions. This, in turn, helps to reduce the volatility in interest rates.

The minimum reserve ratio was not designed, however, to be used to make changes in monetary policy. In other words, it was not used as a variable minimum reserves ratio, and for this reason it was set at a low level. From 1 January 1999 to 17 January 2012 the ratio was 2 per cent of key liquid and relatively liquid liabilities. However, as of 18 January 2012 the ratio was reduced to 1 per cent. This was to help stimulate bank lending and was therefore part of an *active* monetary policy. The ratio has remained at 1 per cent ever since.

The main instrument for keeping the ECB's desired interest rate as the equilibrium rate is open-market operations in government bonds and other recognised assets, mainly in the form of repos. These repo operations are conducted by the national central banks, which must ensure that the repo rate does not rise above the marginal overnight lending rate or fall below the deposit rate.

[1] See http://www.ecb.int/ecb/orga/decisions/govc/html/index.en.html.

- If people think the economy is going to grow faster, the demand for loans will increase as firms seek to increase their investment.

It is very difficult for the central bank to predict what people's expectations will be. Speculation depends considerably on world political events, rumour and 'random shocks'.

If the demand curve shifts very much, and if it is inelastic, then monetary control will be very difficult. Furthermore, the central bank will have to make frequent and sizeable adjustments to interest rates. These fluctuations can be very damaging to business confidence and may discourage long-term investment thereby reducing long-run economic growth.

CASE STUDIES & APPLICATIONS

The ECB ordinarily uses two principal types of open-market operations:

Main refinancing operations (MROs). These are short-term repos with a maturity of one week. They take place weekly and are used to maintain liquidity consistent with the chosen ECB interest rate.

Longer-term refinancing operations (LTROs). These take place monthly and normally have a maturity of three months. Longer maturities are available, but such operations are conducted more irregularly. They are to provide additional longer-term liquidity to banks as required at rates determined by the market, not the ECB.

Non-standard policy measures

The financial crisis put incredible strains on commercial banks in the eurozone and on the real economy. Consequently, monetary operations were gradually modified.

Securities Market Programme (SMP). In May 2010 the ECB began purchases of central government debt in the secondary market (i.e. not directly from governments) as well as purchases both in primary and secondary markets of private-sector debt instruments. Designed to supply liquidity to the ailing banking sector, the SMP saw €214 billion of purchases made by June 2012. These were largely of government bonds issued by countries experiencing financing difficulties, including Portugal, Ireland, Greece and Spain.

However, purchases under the SMP were offset by selling securities elsewhere. This offsetting process, known as **sterilisation,** was designed to leave the overall money supply unchanged.

Three-year long-term refinancing operations (LTROs). LTROs were designed to provide liquidity to financial institutions and stimulate lending. Operations in December 2011 and February 2012 saw €1 trillion drawn on by around 800 banks at a fixed interest rate of 1 per cent.

Targeted long-term refinancing operations (TLTROs). These TLTROs were to provide long-term loans to commercial banks at cheap rates to stimulate lending. The amounts that banks could borrow were linked to the size of eligible outstanding loans to households and non-financial corporations, while the interest rate charged was dependent on the lending history of banks.

The first series of TLTROs was announced in June 2014, a second series in March 2016 (TLTRO-II) and a third series in July 2019 (TLTRO-III). The third TLTRO programme began in September 2019 with loans of up to three years. Rates were cut to as low as –1 per cent. The last of 10 tranches was released in December 2021, taking the amount drawn to €1.5 trillion.

Quantitative easing (QE). In September 2014, the ECB finally announced that it would be commencing quantitative easing through an asset purchases programme (APP), including public-sector bonds and securitised mortgages (see Section 12.2 and Box 13.6). Between March 2015 and December 2018, the ECB made purchases worth €2.6 trillion.

Although purchases halted in December 2018, this proved to be only a temporary pause. With low growth and inflation around only 1 per cent, the ECB announced in September 2019 that QE would restart in November. It committed itself to asset purchases of €20 billion per month for as long as was necessary. However, the COVID-19 pandemic meant that an *additional* programme of asset purchases known as the Pandemic Emergency Purchase Programme (PEPP) was announced. The size of PEPP was subsequently expanded twice, taking total asset purchases to €1850 billion.

The inflation shock of 2021–22 saw the rate of inflation, as in many countries, rise significantly above target. This led to a tightening of monetary policy and an accelerated unwinding of asset purchases. In December 2021, the ECB announced that net asset purchases under PEPP would cease from March 2022. Then, in March 2022, it announced that those under APP would cease that summer. In July 2022, the ECB announced its first increase in interest rates since July 2011.

What are the arguments for and against publishing the minutes of the meetings of the ECB's Governing Council and Executive Board?

Undertake a literature search on what is meant by non-standard monetary policy measures. Write a short summary of your findings.

Definition

Sterilisation Actions taken by a central bank to offset the effects of foreign exchange flows or its own bond transactions so as to leave money supply unchanged.

Pause for thought

Assume that the central bank announces a rise in interest rates and backs this up with open-market operations. What determines the size of the resulting fall in aggregate demand?

There is also the problem of how reliable the *pass-through effect* is from changes in central bank rates to other interest rates in the economy. This may be a particular problem when nominal interest rates are close to zero. To help address this problem following both the financial crisis and the pandemic, central banks pumped large amounts of money into the economy (see Box 13.6) and used targeted long-term repo operations to provide money to banks for long-term loans (see Box 13.5).

BOX 13.6 QUANTITATIVE EASING

Rethinking monetary policy in hard times

Increasing the money supply

As the economies of the world slid into recession in 2008, central banks became more and more worried that the traditional instrument of monetary policy – controlling interest rates – was insufficient to ward off a slump in demand.

Interest rates had been cut at an unprecedented rate and central banks were reaching the end of the road for further cuts. Despite nominal interest rates being close to zero, they were having little effect on aggregate demand. The problem was that there was an acute lack of willingness of banks to lend, and firms and consumers to borrow, as people saw the oncoming recession.

So, what were central banks to do? As we have seen above (see page 391 and Boxes 13.4 and 13.5) the answer was to increase money supply directly through **quantitative easing (QE).** This involves an aggressive version of open-market operations, where the central bank buys up a range of assets, such as long-term government bonds, corporate bonds and securitised mortgage debt. The effect is to pump large amounts of additional cash into the economy in the hope of stimulating demand and, through the process of credit creation, to boost broad money too.

QE in the USA

In the USA, in December 2008, at the same time as the federal funds rate was cut to a range of 0 to 0.25 per cent, the Fed embarked on large-scale quantitative easing. It began buying hundreds of billions of dollars' worth of mortgage-backed securities on the open market and planned also to buy large quantities of long-term government debt.

The result was that considerable quantities of new money were injected into the system. At the conclusion of three rounds of QE in October 2014, the Fed had purchased assets of $2.5 trillion.

QE resumed in March 2020 (QE4) in response to COVID-19, with an open-ended commitment by the Fed. When in March 2022 the Fed conducted its final round of asset purchases, its balance sheet had doubled to almost $9 trillion.

QE in the UK

A similar approach was adopted in the UK. In January 2009, the Bank of England was given powers by the Treasury to buy up to £50 billion of high-quality private-sector assets, such as corporate bonds and commercial paper. The purchases from non-bank financial institutions were with newly created electronic money. But this was only the start.

In March 2009, as the recession deepened, the Chancellor agreed to increase the scale of purchases, so beginning the second and much more significant phase of quantitative easing. These purchases, mainly government bonds (gilts), resulted in a substantial increase in banks' reserves at the Bank of England and hence in the Bank's balance sheet. By July 2012 asset purchases had been made totalling £375 billion. In August 2016, with concerns about the prospects for the UK economy following the vote to leave the European Union, the Bank of England increased asset purchases by a further £70 billion.

Then between March 2020 and December 2021, in response to the COVID-19 pandemic, the Bank increased asset purchases by a further £450 billion, taking its asset purchases to £895 billion.

QE in the eurozone

As we saw in Box 13.5, ECB was more reticent about adopting quantitative easing. But, with the increasing risk of deflation (falling prices), the Asset Purchases Programme (APP) began in March 2015 at a rate of €60 billion of asset purchases per month. This involved purchases of public-sector bonds and securitised mortgages and commercial loans (see Section 12.2).

Initially the scheme was to run to September 2016, with total purchases of up to €1.08 trillion. However, with the inflation rate now only around zero, the end of the scheme was delayed several times. Monthly purchases rose in March 2016 to €80 billion (and including corporate bonds for the first time), before falling back to €60 per month from April 2017 and then to €30 billion from January 2018.

The first round of QE finally came to an end in December 2018, by which point asset purchases had reached

Using monetary policy

It is impossible to use monetary policy as a precise means of controlling aggregate demand. It is especially weak when it is pulling against the expectations of firms and consumers, and when it is implemented too late. However, if the authorities operate a tight monetary policy firmly enough and long enough, they should eventually be able to reduce lending and hence aggregate demand, and with it, inflation. But there will inevitably be time lags and imprecision in the process.

An expansionary monetary policy is even less reliable. If the economy is in recession, no matter how low interest rates are driven, or however much the monetary base is expanded, people cannot be forced to borrow if they do not wish to. Firms will not borrow to invest if they predict a continuing recession.

A particular difficulty in using interest rate reductions to expand the economy arises if the repo rate is nearly zero. As we saw earlier, there is the problem of a weak pass-through effect: the loosening of monetary policy may not be mirrored by an easing of credit conditions. Part of the problem is that (nominal) interest rates cannot be negative (except in the case of overnight deposits by banks in the central bank), for clearly nobody would be willing to lend in

CASE STUDIES & APPLICATIONS

€2.6 trillion. However, a second round of QE began in November 2019 at a rate of €20 billion per month. This was then supplemented from March 2020 by the Pandemic Emergency Purchase Programme (PEPP). Total asset purchases under PEPP were initially set at €750 billion but raised in December 2020 to €1850 billion. Net purchases under PEPP ceased in March 2022, and APP shortly afterwards, by which time the ECB had made total asset purchases of €3.2 trillion.

Transmission mechanisms

Quantitative easing involves directly increasing the amount of narrow money. It can also, indirectly, increase broad money. There are two principal ways in which this can happen.

Asset prices and yields. When non-bank financial companies, including insurance companies and pension funds, sell assets to the central bank, they can use the money to purchase other assets, such as shares. In doing so, this will drive up their prices. This, in turn, reduces the yields on these assets (at a higher price there is less dividend or interest per pound spent on them), which should help to reduce interest rates generally and make the cost of borrowing cheaper for households and firms, so boosting aggregate demand.

Also, for those holding these now more expensive assets there is a positive wealth effect. For instance, households with longer-term saving plans involving securities will now have greater financial wealth. Again, this will boost spending.

Bank lending. Commercial banks will find their reserve balances increase at the central bank as the sellers of assets to the central bank deposit the money in their bank accounts. This will increase the liquidity ratio of banks, which could encourage them to grant more credit.

However, it is all very well increasing the monetary base, but a central bank cannot force banks to lend or people to borrow. That requires confidence. As Figure 11.7 (page 328) shows, bank lending to the non-bank private sector was weak for some time in the early 2010s and significantly below pre-financial crisis levels.

This is not to say that quantitative easing failed in the UK and elsewhere: growth in broad money could have been weaker still. However, it does illustrate the potential danger of this approach if, in the short run, little credit creation takes place. In the equation $MV = PY$, the rise in (narrow) money supply (M) may be largely offset by a fall in the velocity of circulation (V) (see pages 337–8).

On the other hand, there is also the danger that if this policy is conducted for too long, the growth in broad money supply can ultimately prove to be excessive, resulting in inflation rising above the target level.

As it turned out, it was cost-push pressures that would cause central banks to turn off the monetary 'tap' as inflation rates rose significantly in 2021–22. This inflation shock saw a tightening of monetary policy and a period of 'quantitative tightening' – *selling* assets that the central bank had purchased, thereby driving down asset prices and driving up interest rates.

Would it be appropriate to define the policy of quantitative easing as 'monetarist'?

Undertake desktop research, including visiting the websites of major central banks such as the Bank of England, the ECB and the Federal Reserve. Search for materials explaining the process of the quantitative easing (QE) mechanisms by which it is expected to impact on inflation and economic activity and any assessment of its effects where QE has been employed. Briefly summarise your findings.

Definition

Quantitative easing (QE) A deliberate attempt by the central bank to increase the money supply by buying large quantities of securities through open-market operations. These securities could be securitised mortgage and other private-sector debt or government bonds.

these circumstances. Many countries were in this position in the early 2010s. They found themselves caught in what is known as the **liquidity trap**. Despite record low interest rates and high levels of liquidity, borrowing and lending remained low given worries about fiscal austerity and its dampening effects on economic growth.

Forward guidance. One way in which central banks, like the Federal Reserve, the Bank of England and the ECB, attempted to encourage spending following the financial crisis of the late 2000s was by publicly indicating the expected path of future interest rates. By stating that interest rates were likely to remain low for some time, central banks hoped that this *forward guidance* would give economic agents confidence to bring forward their spending.

TC 11 p 61

Despite these problems, changing interest rates can often be quite effective in the medium term. After all, they can be changed very rapidly. There are not

Definition

Liquidity trap When interest rates are at their floor and thus any further increases in money supply will not be spent but merely be held in idle balances as people wait for the economy to recover and/or interest rates to rise.

the time lags of implementation that there are with fiscal policy. Indeed, since the early 1990s, most governments or central banks have used interest rate changes as the major means of keeping aggregate demand and inflation under control. What is more, having an inflation target and using active monetary policy to ensure that inflation is kept at or close to the target sends a very clear message to people that inflation *will* be kept under control. This helps to anchor inflationary expectations – the expected rate of inflation becomes the target rate.

If, however, people no longer believe that central banks will be successful, as was the case in early 2022 with the rapid rise in energy and food prices, inflationary expectations will rise and central banks may have to use quite aggressive policy to bring inflation down, contributing to a contraction of the economy.

Recap

1. Control of the growth in the money supply over the longer term will normally involve governments attempting to restrict the size of public-sector deficits.
2. In the short term, the central bank can use monetary policy to restrict/increase the growth in aggregate demand in one of two major ways: (a) reducing/increasing money supply directly, (b) reducing/increasing the demand for money by raising/lowering interest rates.
3. The money supply can be reduced/increased directly by using open-market operations. This involves the central bank selling more government securities and thereby reducing/increasing banks' reserves. Alternatively, the central bank can reduce/increase the amount of lending or rediscounting it is prepared to do (other than as a last-resort measure).
4. Rather than controlling the monetary base in either of these two ways, the authorities could use funding. This involves increasing/reducing the sale of bonds relative to bills, thereby reducing/increasing banks' liquid assets. Finally, it could operate a system of variable minimum reserve ratios. Increasing these would force banks to cut back the amount of credit they create; reducing them would allow banks to create more credit.
5. Controlling either the monetary base or broad liquidity in the short term, however, is difficult given that central banks are always prepared to provide liquidity to the banks on demand. Even if the authorities are successful in controlling the money supply, there then arises the problem of severe fluctuations in interest rates if the demand for money fluctuates and is relatively inelastic.
6. The normal method of control in most countries involves the central bank influencing interest rates by its operations in the gilt repo and discount markets. The central bank keeps banks short of liquidity and then supplies them with liquidity, largely through gilt repos, at its chosen interest rate (gilt repo rate). This then has a knock-on effect on interest rates throughout the economy.
7. With an inelastic demand for loans, however, changes in interest rates may have to be very large to bring the required changes in monetary growth. High interest rates are politically unpopular and discriminate against those with high borrowing commitments. They also drive up the exchange rate, which can damage exports.
8. Controlling aggregate demand through interest rates is made even more difficult by fluctuations in the demand for money. These fluctuations are made more severe by speculation against changes in interest rates, exchange rates, the rate of inflation, etc.
9. It is impossible to use monetary policy as a precise means of controlling aggregate demand in the short term. Nevertheless, controlling interest rates is a rapid way of responding to changing forecasts, and can be an important signal to markets that inflation will be kept under control, especially when there is a firm target for the rate of inflation.
10. After the global financial crisis and then again in response to the COVID-19 pandemic, many central banks embarked on programmes of quantitative easing, which involved their creating large amounts of new narrow money. This was used to purchase bonds and other assets from financial institutions, thereby increasing banks' reserves and allowing them to increase lending and hence increase broad money through the process of credit creation.

13.4 DEMAND-SIDE POLICY

What will be the effect of attempts by the government to control the level of spending in the economy?

Attitudes towards demand management

Debates over the control of demand have shifted ground somewhat in recent years. There is now less debate over the relative effectiveness of fiscal and monetary policy in influencing aggregate demand. There is general agreement that a *combination* of fiscal and monetary policies will have a more powerful effect on demand than either used separately.

Economists have become increasingly interested in the environment within which policy is made. In this section we analyse debates around the extent to which governments ought to pursue active demand management policies or adhere to a set of policy rules.

Those in the Keynesian tradition prefer discretionary policy – changing policy as circumstances change. TC 3 p 10 Those in the monetarist and new classical tradition prefer to set firm rules (e.g. targets for inflation, public deficits or growth in the money supply) and then stick to them.

The case for rules and policy frameworks

Why should governments commit to rules or design policy frameworks which may involve their giving up control of economic instruments? There are two key arguments against discretionary policy.

Political behaviour. The first concerns the motivation of government. Politicians may attempt to manipulate the economy for their own political purposes – such as the desire to be re-elected. The government, if not constrained by rules, may over-stimulate the economy some time before an election so that growth is strong at election time. After the election, the government strongly dampens the economy to deal with the higher inflation and rising public-sector debt and to create enough slack for another boost in time for the next election.

If the effect of manipulating economic policy is to generate cycles in output, unemployment and inflation that mirror the political cycle, then the result is a **political business cycle**. The theory of the political business cycle suggests that the discretionary policy choices of government can purposefully destabilise the economy. Therefore, it is argued, governments should be made to adopt policy rules.

The manipulation of policy instruments may not be necessarily as systematic or regular in the way that the political business cycle model implies. Nonetheless, the manipulation is intended to court short-term favour with the public and may store up problems for the economy and, in the case of fiscal policy, for the public finances.

It is argued that when politicians behave in this way, fiscal policy may exhibit a **deficit bias**. Because governments are more willing to use their discretion to loosen fiscal policy than they are to tighten fiscal policy, persistent deficits and a rising debt-to-GDP ratio can result. Table 13.1 (on page 376) provides some support for this. Therefore, fiscal rules may be needed to ensure the long-term sustainability of public finances.

TC 11 p 61 It is also argued that politically motivated policymakers can lose *credibility* for sound economic management. This can lead to higher inflationary expectations, uncertainty and lower long-term investment. Firms are likely to increase prices and workers to demand higher wages should government loosen its policy stance to try to boost its popularity. The result could be an *inflation bias*, with rates of inflation higher than they would otherwise be, but with unemployment no lower.

As we saw in Section 12.5, the possibility of inflation bias (see pages 369–70) is one of the principal arguments behind the move in many countries for the delegation of monetary policy to central banks.

Time lags with discretionary policy. Both fiscal and monetary policies can involve long and variable time lags, which can make the policy at best ineffective and at worst destabilising. Taking the measures *before* the problem arises, and thus lessening the problem of lags, is often not a realistic option since forecasting tends to be unreliable. KI 30 p 203

In contrast, by setting and sticking to rules, and then not interfering further, the government can provide a sound monetary framework in which there is maximum freedom for individual initiative and enterprise, and in which firms are not cushioned from market forces and are therefore encouraged to be efficient. By the government setting a target for a steady reduction in the growth of money supply, or a target for the rate of inflation, and then resolutely sticking to it, people's expectations of inflation will be reduced, thereby making the target easier to achieve. TC 11 p 61

This sound and stable monetary environment, with no likelihood of sudden contractionary or expansionary fiscal or monetary policy, will encourage firms to take a longer-term perspective and to plan ahead. This could then lead to increased capital investment and, as we saw in Chapter 9, raise long-term growth rates. TC 11 p 61 TC 15 p 287

The optimum situation is for all the major countries to adhere to mutually consistent rules, so that their economies do not get out of line. This will create more stable exchange rates and provide the climate for world growth (we explore these issues in Chapters 14 and 15).

Advocates of this point of view in the 1970s and 1980s were monetarists and new classical macroeconomists,

Definitions

Political business cycle The theory that governments will engineer an economic contraction, designed to squeeze out inflation, followed by a pre-election boom.

Deficit bias The tendency for frequent fiscal deficits and rising debt-to-GDP ratios because of the reluctance of policy makers to tighten fiscal policy.

but support for the setting of targets was to become widespread. As we have seen, targets are often set for both inflation and fiscal indicators, such as public-sector deficits.

The case for discretion

Keynesians typically reject the argument that rules provide the environment for high and stable growth. Demand, argue Keynesians, is subject to many and sometimes violent shocks, such as changes in expectations, domestic political events (e.g. impending elections), world economic factors (e.g. the global financial crisis), world political events (e.g. the war in Ukraine) and health emergencies (e.g. the COVID-19 pandemic). The resulting shifts in injections or withdrawals cause the economy to deviate from a stable full-employment growth path.

TC 3 p 10

Any change in injections or withdrawals will lead to a cumulative effect on national income via the multiplier, (investment) accelerator and financial accelerator and via changing expectations. These effects take time and interact with each other, and so a process of expansion or contraction can last many months before a turning point is eventually reached.

TC 14 p 249

Since shocks to demand occur at irregular intervals and are of different magnitudes, the economy is likely to experience cycles of irregular duration and of varying intensity. Given that the economy is inherently unstable and is buffeted around by various shocks, Keynesians argue that the government needs actively to intervene to stabilise the economy. Otherwise, the uncertainty caused by unpredictable fluctuations will be very damaging to investment and hence to long-term economic growth (quite apart from the short-term effects of recessions on output and employment).

TC 3 p 10

If demand fluctuates in the way Keynesians claim, and if the policy of having a money supply or inflation rule is adhered to, interest rates must fluctuate. But excessive fluctuations in interest rates will discourage long-term business planning and investment. What is more, the government may find it difficult to keep to its targets. This too may cause uncertainty and instability.

Problems with rules and targets

The global financial crisis and pandemic meant that many economies suspended fiscal rules twice in a little over a decade. Even when a framework has been in force for some time, it may cease to be appropriate. Persistently low inflation rates during the 2010s contributed to the ECB adopting a symmetric inflation target of 2 per cent over the medium term (see pages 394 and 474).

Then there is the question of whether success in achieving the target will bring success in achieving other macroeconomic objectives, such as low unemployment and stable economic growth. The problem is that something called **Goodhart's Law** is likely to apply. The law, named after Charles Goodhart of the LSE and formerly of the Bank of England, states that attempts to control an *indicator* of a problem may, as a result, make it cease to be a good indicator of the problem.

KEY IDEA 35

Goodhart's Law. **Controlling a symptom (i.e. an indicator) of a problem will not cure the problem. Instead, the indicator will merely cease to be a good indicator of the problem.**

Targeting inflation may make it become a poor indicator of the state of the economy (see Section 12.5). If people believe that the central bank will be successful in achieving its inflation target, then those expectations will feed into their inflationary expectations, and not surprisingly the target will be met. But that target rate of inflation may now be consistent with both a buoyant and a depressed economy. In other words, the Phillips curve may become *horizontal*.

The implication of Goodhart's Law is that, in achieving their inflation target, policy makers may not be tackling the much more serious problem of creating stable economic growth and an environment that will therefore encourage long-term investment.

In extreme cases, where growth may be slowing, such as in 2008 in the face of the global financial crisis, or just recovering, such as in 2021–22 as countries began to ease pandemic restrictions, economies may face cost-push factors, causing inflation to rise. Strictly adhering to an inflation target in these circumstances will demand *higher* interest rates, which could further restrict growth. When economies experience cost-push inflation, inflation-targeting central banks may be confronted with an inflation–output stabilisation trade-off.

KI 31 p 206

Pause for thought

Under what circumstances would adherence to money supply targets lead to (a) more stable interest rates and (b) less stable interest rates than pursuing discretionary demand management policy?

Definition

Goodhart's Law Controlling a symptom of a problem, or only part of the problem, will not cure the problem: it will simply mean that the part that is being controlled now becomes a poor indicator of the problem.

Central banks and a Taylor rule

Given the potential problems in adhering to simple inflation rate targets, many economists have advocated the use of a **Taylor rule**,[1] rather than a simple inflation target. A Taylor rule takes two objectives into account – (1) inflation and (2) either real national income or unemployment – and seeks to get the optimum degree of stability of the two. The degree of importance attached to each of the two objectives can be decided by the government or central bank. The central bank adjusts interest rates when either the rate of inflation diverges from its target or the level of real national income (or unemployment) diverges from its sustainable (or equilibrium) level.

Take the case where the rate of inflation is above its target level. The central bank following a Taylor rule will raise the rate of interest. It knows, however, that this will reduce real national income below the level at which it would otherwise have been. This, therefore, limits the amount that the central bank is prepared to raise the rate of interest. The more weight it attaches to stabilising inflation, the more it will raise the rate of interest. The more weight it attaches to achieving stable growth in real national income, the less it will raise the rate of interest.

KI 31 p 206

Thus the central bank has to trade off inflation stability against stability in economic growth. As in the early 2020s, this trade-off can be large when there are significant cost-push factors affecting the rate of inflation. The weights attached to these two objectives are therefore especially important in such circumstances.

Pause for thought

If people believe that the central bank will be successful in keeping inflation on target, does it matter whether a simple inflation rule or a Taylor rule is used? Explain.

[1]Named after John Taylor, from Stanford University, who proposed that for every 1 per cent that inflation rises above its target level, real interest rates should be raised by 0.5 percentage points (i.e. nominal rates should be raised by 1.5 percentage points) and that for every 1 per cent that real national income rises above its potential level, real interest rates should be raised by 0.5 percentage points (i.e. nominal rates should be raised by 0.5 percentage points).

Conclusions

The following factors provide us with a framework to help analyse the relative merits of rules or discretion:

- The confidence of people in the effectiveness of either discretionary policies or rules: the greater the confidence, the more successful is either policy likely to be.
- The degree of self-stabilisation of the economy (in the case of rules), or conversely the degree of inherent instability of the economy (in the case of discretion).
- The size and frequency of exogenous shocks to demand: the greater they are, the greater the case for discretionary policy.
- In the case of rules, the ability and determination of governments to stick to the rules and the belief by the public that they will be effective.
- In the case of discretionary policy, the ability of governments to adopt and execute policies of the correct magnitude, the speed with which such policies can be implemented and the accuracy of forecasting.

Case Study 13.20 on the student website looks at the history of fiscal and monetary policies in the UK from the 1950s to the current day. It illustrates the use of both rules and discretion and how the debates about policy shifted with historical events.

Definition

Taylor rule A rule adopted by a central bank for setting the rate of interest. It will raise the interest rate if (a) inflation is above target or (b) real national income is above the sustainable level (or unemployment is below the equilibrium rate). The rule states how much interest rates will be changed in each case.

Recap

1. The case against discretionary policy is that it involves unpredictable time lags that can make the policy destabilising. Also, the government may *ignore* the long-run adverse consequences of policies designed for short-run political gain.
2. The case in favour of rules is that they help to reduce deficit bias and inflationary bias and help create a stable environment for investment and growth.
3. The case against sticking to money supply or inflation rules is that they may cause severe fluctuations in interest rates and thus create a less stable economic environment for business planning. Given the changing economic environment in which we live, rules adopted in the past may no longer be suitable for the present.
4. Although perfect fine-tuning may not be possible, Keynesians argue that the government must have the discretion to change its policy as circumstances demand.

13.5 SUPPLY-SIDE POLICY

How might the government attempt to control the level of output and employment directly?

TC 7 p 26

Supply-side policies, as the name suggests, focus on aggregate supply. If successful, they will shift the aggregate supply curve to the right, thus increasing output for any given level of prices (or reducing the price level for any given level of output).

In Chapter 10 we saw that aggregate supply depends on an economy's factors of production, including labour and capital. Therefore, policies which seek to influence the *quantities* of the factors employed or their *productivity* constitute supply-side policies. If successful, they will not only increase potential output but raise the rate of long-term economic growth and positively impact on general living standards. They are also likely to affect employment and unemployment and put downward pressure on prices.

TC 15 p 287

Supply-side policies are commonly grouped under two general types: *market-orientated* and *interventionist*. Market-orientated policies focus on ways of 'freeing up' the market, such as encouraging private enterprise, risk-taking and competition: policies that provide incentives for innovation, hard work and productivity. Interventionist policies focus on means of counteracting the deficiencies of the free market and typically involve government expenditure on infrastructure and training and financial support for investment.

TC 6 p 24

However, some supply-side policies may draw on elements of both types, for instance, by providing financial support (interventionist) through the use of tax reliefs (market-orientated). Such policies are sometimes described as 'third-way' supply-side policies.

Market-orientated supply-side policies

Radical market-orientated supply-side policies were first adopted in the early 1980s by the Thatcher government in the UK and the Reagan administration in the USA, but were subsequently copied by other right and centre-right governments around the world. The essence of these policies is to encourage and reward individual enterprise and initiative, and to reduce the role of government; to put more reliance on market forces and competition, and less on government intervention and regulation.

Reducing government expenditure

The desire by many governments to cut government expenditure is not just to reduce public-sector deficits; it is also an essential ingredient of their supply-side strategy.

The public sector is portrayed by some as more bureaucratic and less efficient than the private sector. What is more, it is claimed that a growing proportion of public money has been spent on administration and other 'non-productive' activities, rather than on the direct provision of goods and services.

Two things are needed, it is argued: (a) a more efficient use of resources within the public sector and (b) a reduction in the size of the public sector. This would allow private investment to increase with no overall rise in aggregate demand. Thus the supply-side benefits of higher investment could be achieved without the demand-side costs of higher inflation.

In practice, governments have found it very difficult to cut their expenditure relative to GDP. However, many countries were faced with trying to do this after interventions in response first to the financial crisis and later to the COVID-19 pandemic saw public expenditure (see Figure 13.6) and, in turn, budget deficits increase significantly. Concerns around the sustainability of the public finances can therefore result in often difficult fiscal choices needing to be made.

Tax cuts

The imposition of taxation can distort a variety of choices that individuals make. Changes to the rates of taxation can lead individuals to substitute one activity for another. Three examples that are commonly referred to in the context of aggregate supply are:

TC 6 p 24

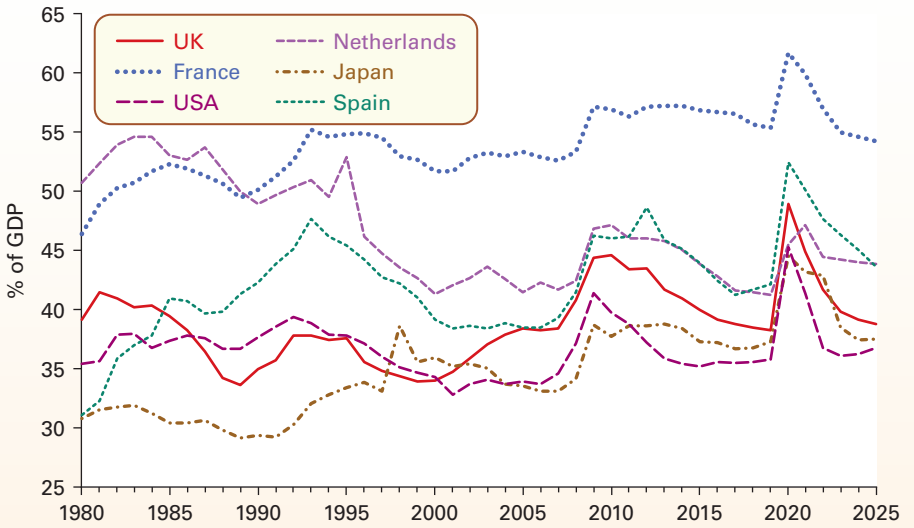
- taxation of income and its impact on labour supply (including hours worked and choice of occupation);
- taxation of interest income earned on financial products (savings) and its impact on the funds available for investment;
- taxation of firms' profits and its impact on capital expenditure by firms.

Over time, many countries have witnessed a decline in the marginal rates of taxation associated with each of these cases. Here we consider the UK.

In 1979, the basic rate of income tax in the UK was 33 per cent, with higher rates rising to 83 per cent. By 2008 the standard rate was 20 per cent and the higher rate was 40 per cent. From 2010, an additional 50 per cent tax rate was implemented for those earning in excess of £150 000, largely as a means of plugging the deficit in the public finances. This was subsequently reduced to 45 per cent from 2013. However, in November 2022 it was announced that from April 2023 the threshold for the additional rate would be reduced to £125 140.

Similar reductions in rates of tax on business profits have been designed to encourage investment. Reductions in corporation tax (the tax on business

Figure 13.6 General government expenditure (% of GDP)



Note: Data from 2022 based on forecasts

Source: Based on data from *World Economic Outlook Database* (IMF) and *AMECO Database* (European Commission) for USA 1980–2000

profits) have increased after-tax profits, leaving more funds for ploughing back into investment, as well as increasing the after-tax return on investment. In 1983 the main rate of corporation tax in the UK stood at 52 per cent. A series of reductions then took place, with the main rate falling to 19 per cent by 2017. The pandemic brought an end to further cuts and the rate was increased to 25 per cent from April 2023.

Governments have also looked to increase investment allowances or R&D expenditure credits. Investment allowances enable firms to offset the cost of investment against pre-tax profit, thereby reducing their tax liability. R&D expenditure credits operate by providing firms with cash payments for a proportion of their R&D expenditure which, although subject to tax, nonetheless increase their net profit.

Successive governments have applied such R&D incentives. For example, the UK operates a Research and Development Expenditure Credit (RDEC). From April 2023, the rate increased from 13 per cent to 20 per cent.

Additionally, all firms are eligible for a 100 per cent R&D relief from their yearly profit, with small and medium-sized enterprises (SMEs) able to claim additional relief. However, this additional relief for SMEs was reduced from 130 per cent to 86 per cent from April 2023.

Since April 2013, firms have been subject to a lower rate of corporation tax on profits earned from inventions they have patented and certain other innovations. The idea is that firms will be provided with financial support to innovate where this results in their acquiring patents. Patents provide protection for intellectual property rights. Firms are liable to corporation tax on the profits attributable to qualifying patents at a reduced rate of 10 per cent.

The argument for reducing tax rates on incomes and profits is that it contributes to higher real GDP. Specifically, it contributes to there being more labour hours supplied, more moneys invested with financial institutions and more capital expenditure by firms than would otherwise be the case. In other words, there is a substitution effect (see page 186) inducing more of these beneficial activities. In the case of labour, people are encouraged to substitute work for leisure.

TC 2 p 9

However, in each case there is a counteracting incentive from the tax cut: an income effect (see page 186). This is because the higher returns to working, saving and undertaking capital expenditure mean that less of each activity needs to be undertaken to generate the same income flow as before.

Because economic theory offers no firm conclusions as to the benefit of tax cuts, economists and policy makers often look to empirical evidence for guidance. The evidence on whether people will be prepared to work longer hours, for instance, suggests that the substitution and income effects just about cancel each other out. Anyway, for many people there is no such choice in the short run. There is no chance of doing overtime or working a shorter week. In the long run, there may be some flexibility in that people can change jobs.

Reducing the power of labour

The argument here is that, if labour costs to employers are reduced, their profits will probably rise. This could encourage and enable more investment and hence economic growth. If the monopoly power of labour is reduced, then cost-push inflation will also be reduced.

The Thatcher government in the 1980s took a number of measures to curtail the power of unions. These included the right of employees not to join unions, preventing workers taking industrial action other than against their direct employers and enforced secret ballots on strike proposals. It set a lead in resisting strikes in the public sector.

As labour markets have become more flexible, with increased part-time working and short-term and zero-hours contracts, and as the process of globalisation has exposed more companies to international competition, so this has further eroded the power of labour in many sectors of the economy.

Reducing welfare

New classical economists claim that a major cause of unemployment is the small difference between the welfare benefits of the unemployed and the take-home pay of the employed. This causes voluntary unemployment (i.e. frictional unemployment). People are caught in a 'poverty trap': if they take a job, they lose their benefits (see pages 188–9).

A dramatic solution to this problem would be to cut unemployment benefits. A major problem with this approach, however, is that with changing requirements for labour skills, many of the redundant workers from the older industries are simply not qualified for new jobs that are created. What is more, the longer people are unemployed, the more demoralised they become. Employers would probably be prepared to pay only very low wages to such workers. To persuade these unemployed people to take low paid jobs, the welfare benefits would have to be slashed. A 'market' solution to the problem, therefore, may be a very cruel solution. A fairer one would be an interventionist policy: a policy of retraining labour.

KI 24 p 185

Another alternative is to make the payment of unemployment benefits conditional on the recipient making a concerted effort to find a job. Unemployed workers claiming Universal Credit (or the older scheme, Jobseeker's allowance), must be available for and actively seeking work, and must complete a 'Claimant Commitment', which sets out the types of work the person is willing to do, and the plan to find work. Payment can be refused if the claimant refuses to accept the offer of a job.

TC 6 p 24

Policies to encourage competition

If the government can encourage more competition, this should have the effect of increasing national output and reducing inflation. Five major types of policy have been pursued under this heading.

Privatisation. If privatisation simply involves the transfer of a natural monopoly to private hands (e.g. the water companies), the scope for increased competition is limited. However, where there is genuine scope for increased competition (e.g. in the supply of gas and electricity), privatisation can lead to increased efficiency, more consumer choice and lower prices. There may still be a problem of oligopolistic collusion, however, and thus privatised industries are monitored and regulated with the aim of making them genuinely competitive.

Alternatively, privatisation can involve the introduction of private services into the public sector (e.g. private contractors providing cleaning services in hospitals, or refuse collection for local authorities). Private contractors may compete against each other for the franchise. This may well lower the cost of provision of these services, but the quality of provision may also suffer unless closely monitored. The effects on unemployment are uncertain. Private contractors may offer lower wages and thus may use more labour. But if they are trying to supply the service at minimum cost, they may employ less labour.

Deregulation. This involves the removal of monopoly rights: again, largely in the public sector. The deregulation of the UK bus industry, opening it up to private operators, is a good example of this initiative.

Introducing market relationships into the public sector. This is where the government tries to get different departments or elements within a particular part of the public sector to 'trade' with each other, so as to encourage competition and efficiency.

One example is the UK's National Health Service. In 2003, the government introduced a system of 'foundation trusts'. Hospitals can apply for foundation trust status. If successful, they are given much greater financial autonomy in terms of purchasing, employment and investment decisions. NHS Improvement is responsible for overseeing foundation trusts and NHS trusts.

Critics argue that funds have been diverted to foundation hospitals away from the less well-performing hospitals where greater funding could help that performance.

As far as general practice is concerned, groups of GP practices are formed into Clinical Commissioning Groups (CCGs). These are responsible for arranging most of the NHS services within their boundaries. A key principle of the system is to give GPs a choice of 'providers' with the hope of reducing costs and driving up standards.

Public–private partnerships. Public–private partnerships (PPPs) are a way of funding public expenditure with

private capital. In the UK the *Private Finance Initiative* (PFI), as it was known, began in 1992. The PFI meant that a private company, after a competitive tender, would be contracted by a government department or local authority to finance and build a project, such as a new road or a prison. The government then pays the company to maintain and/or run it, or simply rents the assets from the company. The public sector thus becomes a purchaser of services rather than a direct provider itself.

Critics claim that PFI projects have resulted in low quality of provision and that cost control has often been poor, resulting in a higher burden for the taxpayer in the long term. What is more, many of the projects have turned out to be highly profitable for the private provider, suggesting that the terms of the original contracts were too lax.

Following these criticisms, in 2012 the UK government introduced a revised form of PFI (PF2) where the public sector would take stakes of up to 49 per cent in new projects. These would no longer include 'soft services', such as cleaning and catering. However, following the collapse in January 2018 of Carillion, one of the major private-sector firms involved in PPPs, criticsms of PFI intensified. In response, the Chancellor, in his October 2018 Budget, announced that there would be no further private finance projects, although existing agreements under PFI and PF2 would be honoured. The benefits and costs of PFI and PF2 are explored in Case 13.8 on the student website.

Free trade and capital movements. The opening up of international trade and investment is central to a market-orientated supply-side policy. One of the first measures of the Thatcher government (in October 1979) was to remove all controls on the purchase and sale of foreign currencies, thereby permitting the free inflow and outflow of capital, both long term and short term. Most other industrialised countries also removed or relaxed exchange controls during the 1980s and early 1990s.

The Single European Act of 1987, which came into force in 1993, was another example of international liberalisation. As we shall see in Section 14.6, it created a 'single market' in the EU: a market without barriers to the movement of goods, services, capital and labour.

Critics have claimed that, in the short term, industries may be forced to close by the competition from cheaper imported products, with a major impact on employment in the areas affected. A major plank of the Trump presidency was 'putting America first'. This involved a move away from free trade and giving specific protection to US industries, such as vehicles and steel. We examine the arguments for and against protection in Sections 14.2 and 14.3, while in Section 14.7 we analyse the issue of openness to trade in light of the UK's departure from the EU.

Interventionist supply-side policies

The basis of the case for government intervention is market failure. In particular, in the context of growth of potential output, the free market is likely to provide too little research and development, training and investment. TC 7 p 26

There are potentially large external benefits from research and development. Firms investing in developing and improving products, and especially firms engaged in more general scientific research, may produce results that provide benefits to many other firms. KI 28 p 196

Thus the *social* rate of return on investment may be much higher than the private rate of return. Investment that is privately unprofitable for a firm may therefore still be economically desirable for the nation.

Similarly, investment in training may continue yielding benefits to society that are lost to the firms providing the training when the workers leave.

Investment often involves risks. Firms may be unwilling to take those risks, since the costs of possible failure may be too high. When looked at nationally, however, the benefits of investment might well have substantially outweighed the costs, and thus it would have been socially desirable for firms to have taken the risk. Successes would have outweighed failures. KI 15 p 80

Even when firms do wish to make such investments, they may find difficulties in raising finance. Banks may be unwilling to lend – a problem that increased after the credit crunch. Alternatively, if firms rely on raising finance by the issue of new shares, this makes them very dependent on the stock market performance of their shares. This depends largely on current profitability and expected profitability in the near future, not on *long-term* profitability. Similarly, the fear of takeovers may make managers overconcerned to keep shareholders happy, further encouraging 'short-termism'.

Types of interventionist supply-side policy

Nationalisation. This is the most extreme form of intervention, and one that most countries had tended to reject, given a worldwide trend for privatisation. Nevertheless, many countries had always stopped short of privatising certain key transport and power industries, such as the railways and electricity generation. Having these industries under public ownership may result in higher investment than if they were under private ownership.

Nationalisation may also be a suitable solution for rescuing vital industries suffering extreme market turbulence. This was the case in 2008 with many banks. With the credit crunch and the over-exposure to risky investments in securitised sub-prime debt, inadequate levels of capital, declining confidence and plummeting share prices, several banks were taken into full or

partial public ownership. In the UK, Northern Rock and Bradford & Bingley were fully nationalised, while the government took a majority shareholding in the Royal Bank of Scotland and Lloyds Banking Group.

Direct provision. Improvements in infrastructure, such as a better motorway system, can be of direct benefit to industry. Alternatively, the government could provide factories or equipment to specific firms. The financial crisis and the pandemic raised the spectre of scarring effects: a collapse in current economic activity having adverse consequences for *future* economic outcomes. The IMF, OECD and other international organisations began calling for greater international expenditure on infrastructure, including investments to reduce carbon emissions and improving biodiversity, to increase both aggregate demand and potential output.

Funding research and development. To increase a country's research and development, the government could fund universities or other research institutes through various grants, perhaps allocated by research councils. Alternatively, it could provide grants or tax relief to private firms to carry out R&D.

As we saw above, the UK uses the tax system to encourage research and development (R&D). Despite this, as Figure 13.7 demonstrates, UK gross expenditure on research and development as a percentage of GDP has been lower than that of its main economic rivals.

TC 6 p 24

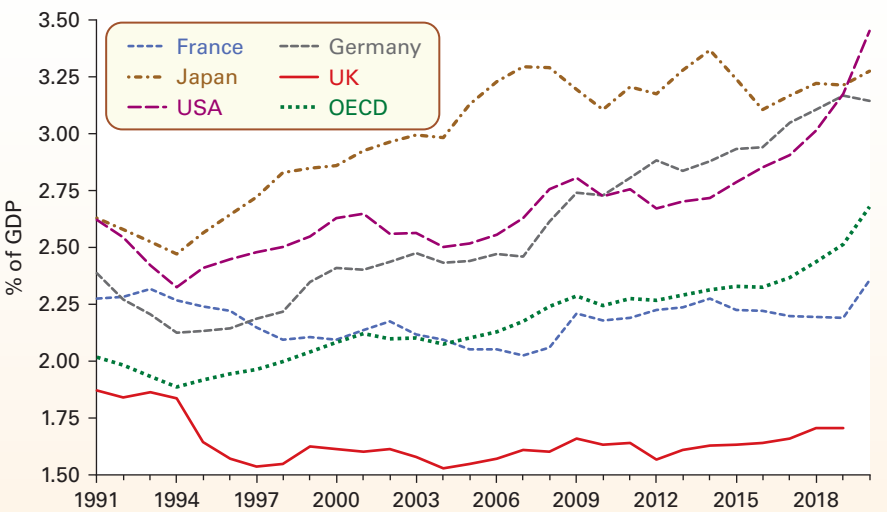
Lower R&D has contributed to a productivity gap between the UK and other advanced economies (see Box 10.2). This productivity gap is a drag on the UK's long-term economic growth. The UK's poor R&D record has occurred even though a sizeable number of UK-based companies are regularly among the world's largest R&D spending companies. In part, this reflects the limited R&D expenditure by government. But, it also reflects the generally lower R&D intensity across the private sector. In other words, total R&D expenditure by British firms has often been low *relative* to the income generated by sales.

Training and education. The government may set up training schemes, or encourage educational institutions to make their courses more vocationally relevant, or introduce new vocational qualifications (such as the GNVQs, NVQs and foundation degrees in the UK). Alternatively, the government can provide grants or tax relief to firms which themselves provide training schemes. Alternative approaches to training in the UK, Germany, France and the USA are examined in Case Study 13.21 on the student website.

Advice, information and collaboration. The government may engage in discussions with private firms in order to find ways to improve efficiency and innovation. It may bring firms together to exchange information and create a climate of greater certainty, or it may bring firms and unions together to try to create greater industrial harmony. It can provide various information services to firms: technical assistance, the results of public research, information on markets, etc.

Local Enterprise Partnerships (LEPs) are an example of partnerships created between local government and businesses in England. LEPs began in 2011 with the aim of promoting local economic development. These partnerships were designed to facilitate the flow of

Figure 13.7 Gross expenditure on R&D as a percentage of GDP



Note: OECD = 38 members (as of 25 May 2021) of the Organisation for Economic Co-operation and Development

Source: *Gross Domestic Spending on R&D* (OECD, 2022)

information between public and private organisations and help them work with local government and other partners to shape local strategic economic objectives.

Regional and urban policy. Certain parts of a country may suffer from lower economic growth and higher unemployment than others. These regional and urban problems normally result from structural problems – the main one being the decline of certain industries, such as mining or heavy manufacturing industry, which had previously been concentrated in those areas.

A response to this problem is for governments or local authorities to provide support for such areas. This can come in various forms, such as giving grants or subsidies for the creation of jobs; expenditure on infrastructure, such as roads and communications, technical colleges, community facilities and housing; the siting of government offices in such areas; the setting up of government agencies to provide advice and support for incoming firms or local community action groups.

In England, 48 Enterprise Zones were established between 2012 and 2016/17. These are specific locations where firms can benefit from reduced planning restrictions, tax breaks and improved infrastructure. By encouraging the *clustering* of specific business types, the hope is that they can mutually benefit from external economies of scale such as co-operation and technological developments.

'Levelling up' became an important element of UK government economic policy following the 2019 election. Its general goals included regenerating towns, helping 'city regions' become globally competitive, and making local growth policies more co-ordinated and coherent. A series of initiatives followed. These included:

- Establishing 'freeports', where raw materials and parts could be imported free of tariffs and made into finished or semi-finished products before being either exported or 'imported' into the country after applying appropriate tariffs.
- A national infrastructure bank to be set up in Leeds and a second Treasury HQ in Darlington.
- A series of levelling up and community investment programmes, such as the £4.8 billion 'Levelling Up Fund' intended for investment in local infrastructure and regeneration.

Pause for thought

How might a new classical economist criticise these various forms of interventionist supply-side policy?

The link between demand-side and supply-side policies

While we may categorise policies as 'demand-side' or 'supply-side' according to their focus, the reality is that policies often have *both* demand-side and supply-side effects. For example, many supply-side policies involve increased government expenditure, whether on retraining schemes, on research and development projects, or on industrial relocation. They will therefore cause a rise in aggregate demand (unless accompanied by a rise in taxes). Similarly, supply-side policies of tax cuts designed to increase incentives will increase aggregate demand (unless accompanied by a cut in government expenditure). It is thus important to consider the consequences for demand when planning various supply-side policies.

Likewise, demand management policies often have supply-side effects. If a cut in interest rates boosts investment, there will be a multiplied rise in national income: a demand-side effect. But that rise in investment will also create increased productive capacity: a supply-side effect.

Recap

1. Market-orientated supply-side policies aim to increase the rate of growth of aggregate supply and reduce the rate of unemployment by encouraging private enterprise and the freer play of market forces.
2. Reducing government expenditure as a proportion of GDP is a major element of such policies.
3. Tax cuts can be used to encourage more people to take up jobs, to work longer hours and to work more enthusiastically. The effects of tax cuts will depend on how people respond to incentives.
4. Reducing the power of trade unions and a reduction in welfare benefits, especially those related to unemployment, may force workers to accept jobs at lower wage rates, thereby decreasing equilibrium unemployment.
5. Other examples of market-orientated supply-side policy include privatisation, competitive tendering for public-sector contracts, deregulation, the Private Finance Initiative, free trade and free movement of capital.
6. Interventionist supply-side policy can take the form of nationalisation, grants for investment and research and development, advice and persuasion, the direct provision of infrastructure and the provision, funding or encouragement of various training schemes.
7. Supply-side policies often have demand-side effects, and demand-side policies often have supply-side effects. It is important for governments to take these secondary effects into account when working out their economic strategy.

BOX 13.7 PUBLIC FUNDING OF APPRENTICESHIPS

Reforms to apprenticeships in England

Current public policy in relation to apprenticeships can arguably be traced back to the launch in 1994 of *Modern apprenticeships*. These were aimed at reversing the decline in apprenticeship numbers which had fallen from 243 700 in 1966 to just 53 000 in 1990. Traditionally, apprenticeships in the UK were developed and run by employers. The new approach involved far greater levels of government intervention with subsidy payments available to help finance the costs of training that met certain criteria.

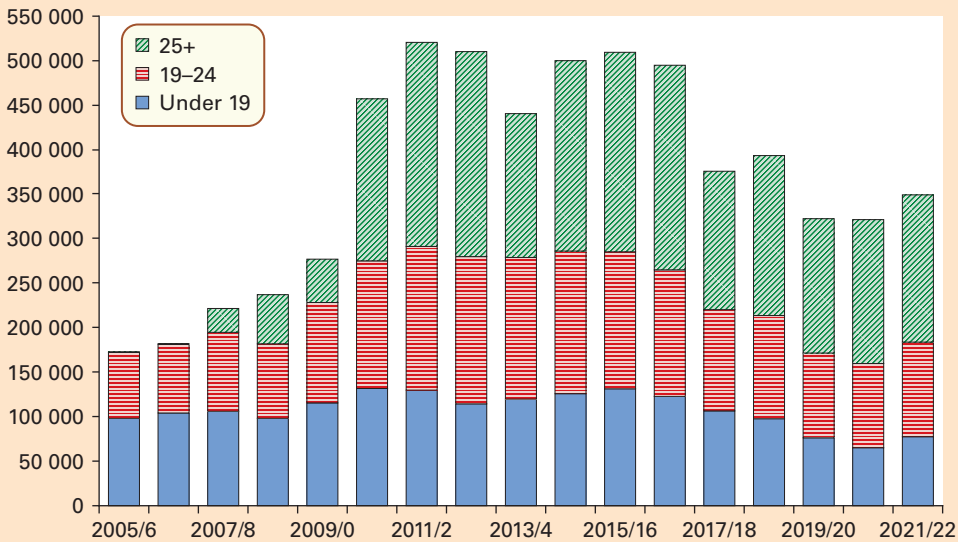
Subsequent reviews of training policy carried by Lord Leitch (2006),[1] Professor Alison Wolf (2011),[2] Dolphin and Lanning (2011),[3] Doug Richard (2012)[4] and Lord Sainsbury (2016)[5] have all agreed on the need for further reform and, above all, to increase the number of publicly funded apprenticeships.

The case for expanding provision

It is argued that expanding the provision of apprenticeships could help reduce youth unemployment rates. Countries with more extensive apprenticeship programmes, such as Germany, Austria and Norway, have much lower rates of youth unemployment than the UK (see Table 12.2 on page 345). Other benefits include, a smoother transition into work for young people not attending university, encouraging employers to engage more fully with training and higher wages for employees who complete low-level vocational qualifications as part of an apprenticeship compared with people who complete the same qualifications outside of an apprenticeship.

Various governments have responded to this recommendation. As the chart shows, England experienced a steady expansion in the number of starts on government-funded apprenticeships between 2006/7 and 2009/10, followed by a more significant expansion in the early 2010s, before then levelling off at around 500 000 per annum. While each age group has seen an increase in the number of apprenticeships, the growth in the number of learners aged 25 and over starting apprenticeships has been particularly rapid.

Total government-funded apprenticeship starts, England (academic years August to July)



Source: Based on data from *Apprenticeships and traineeships* (Department for Education, 25 November 2022)

[1]Lord Leitch, *Prosperity for all in the global economy - world class skills: final report*, Leitch Review of Skills (HM Treasury, December 2006).

[2]Alison Wolf, *Review of vocational education: the Wolf report*, Department for Business, Education & Skills and Department for Education (March 2011).

[3]Tony Dolphin and Tess Lanning, *Rethinking Apprenticeships*, IPPR (November 2011).

[4]Doug Richard, *The Richard Review of Apprenticeships*, GOV.UK (November 2012).

[5]David Sainsbury, *Report of the Independent Panel on Technical Education*, Crown Copyright (2016).

The case for reforming provision

There have been four main arguments for reforming the provision of apprenticeships.

The relevance of the training provided. In traditional apprenticeships, employers play a key role in the design of the schemes, which ensures the training is relevant and highly valued. However, publicly funded apprenticeships have to meet external requirements, such as an Apprenticeship Framework. These set minimum standards detailing both the training and qualifications the employees should receive. The content, design and assessment of most Frameworks were determined by Sector Skills Councils, training providers and qualification bodies. Most employers played a very limited role in the whole process. This often led to the development of skills that were not highly valued by employers.

The level of the training provided. Apprenticeships can be taken at three different levels: *Intermediate* (Level 2, equivalent to 5 GCSEs), *Advanced* (Level 3, equivalent to 2 A-levels) and *Higher* (Levels 4–7, equivalent to undergraduate and postgraduate level). Because the level of government funding available to training providers was partly dependent on key measures of performance, including the proportion of trainees who passed and completed the courses, training providers have an incentive to focus on supplying lower-level intermediate apprenticeships: i.e. the qualifications trainees were least likely to fail. Hence, up to 2015/16, approximately two-thirds of apprenticeship starts were at Intermediate Level, with less than 5 per cent at the Higher Level.

The short duration of many of the schemes. This was a particular issue during the 2010/12 period, when many were completed in less than a year. In some cases, Intermediate Level apprenticeships lasted less than 12 weeks. They typically last two years in most European countries.

The methods of assessment. Assessment under an Apprenticeship Framework involved the trainee completing numerous qualifications, of a short duration, on a continual basis throughout the programme. There was no end-point assessment to make sure the trainee could pull together all of the skills they had learned and demonstrate the competences required for a skilled job.

Recent reforms

The UK government has introduced a number of reforms in response to these criticisms.

In 2017 the government launched the Institute for Apprenticeships and Technical Education. It is responsible for supporting employer groups, known as 'trailblazers', in the development of apprenticeships and in ensuring their quality, including the monitoring of independent assessments of apprentices at the end of their training. The latter involves working with the Office for Qualifications and Examinations Regulation (Ofqual), and the Office for Students (OfS) for integrated higher and degree apprenticeships.

Apprenticeships Standards from 2020 require apprenticeships to (a) be in a skilled occupation; (b) provide substantial and sustained training, lasting a minimum of 12 months, with at least 20 per cent of the training taking place off-the-job; (c) develop transferable skills, including those in maths and English; (d) lead to full competency in an occupation; and (e) provide training that allows employees to apply for professional recognition where it exists.

As part of the government's strategy to support apprenticeships, the Apprenticeship Levy was introduced from April 2017. The apprenticeship levy is a payment that large employers have to make into an online service account. It is calculated as 0.5 per cent of that part of the employer's annual wage bill that exceeds £3 million. Employers in England can then use funds in their apprenticeship accounts to pay for apprenticeship training and assessment with the payment. The amount is aligned to the equivalent educational level of the apprenticeship. However, following the introduction of the Apprenticeship Levy, apprenticeship numbers fell (see chart).

Some argue that the levy is too complex, can be difficult for employers to access, or is seen by employers as another tax. Despite the fall in apprenticeships until 2021/22, the average proportion at the higher level in the two years to 2021/22 was now 30 per cent.

During the COVID-19 pandemic, with particular concerns raised about its impact on the human capital of younger people, the government introduced incentive payments for hiring apprentices in England. These were initially payments of £2000 for the hiring of apprentices aged 16 to 24 and £1500 for those aged 25 and over. The scheme ran from August 2020 to March 2021. This period was subsequently extended until September 2021, with employers receiving an incentive payment of £3000 for hiring apprentices of any age.

The government subsidises course fees for qualifications taken by a trainee as part of an apprenticeship. However, course fees only represent one element of the costs of training a worker. What are the other economic costs?

Research how training taxes/levies operate in at least two other countries. Compare and contrast the schemes with the Apprenticeship Levy in England.

QUESTIONS

1. Does it matter if a country has a large national debt as a proportion of its national income?
2. If the government is running a budget deficit, does this mean that national income will increase?
3. Of what significance are primary deficits or surpluses for the dynamics of a government's debt-to-GDP ratio?
4. How does the size of (a) the budget deficit and (b) public-sector debt vary with the course of the business cycle?
5. How would the withdrawals curve shift in each of the following cases? (a) A reduction in the basic rate of tax. (b) An increase in personal allowances.
6. Under what circumstances is a rise in taxes likely to have a disincentive effect?
7. What factors determine the effectiveness of discretionary fiscal policy?
8. Give some examples of changes in one injection or withdrawal that can affect others.
9. Why is it difficult to use fiscal policy to 'fine-tune' the economy?
10. If the government buys back £1 million of maturing bonds from the general public and then, keeping the total amount of its borrowing the same, raises £1 million by selling bills to banks, what will happen to the money supply?
11. Assume that a bank has the simplified balance sheet shown in the following table, and is operating at its desired liquidity ratio. Now assume that the central bank repurchases £5 million of government bonds on the open market. Assume that the people who sell the bonds all have their accounts with this bank.

Liabilities	£m	Assets	£m
Deposits	100	Balances with central bank	10
		Advances	90
	100		100

    a. Draw up the new balance sheet directly after the purchase of the bonds.
    b. Now draw up the eventual balance sheet after all credit creation has taken place.
    c. Would there be a similar effect if the central bank rediscounted £5 billion of bills?
    d. How would such open-market operations affect the rate of interest?
12. What effect would a substantial increase in the sale of government bonds and bills have on interest rates?
13. Why would it be difficult for a central bank to predict the precise effect on money supply of open-market operations?
14. Imagine you were called in by the government to advise on whether it should adopt a policy of targeting the money supply. What advice would you give and how would you justify the advice?
15. Imagine you were called in by the government to advise on whether it should attempt to prevent cyclical fluctuations by the use of fiscal policy. What advice would you give and how would you justify the advice?
16. Is there a compromise between purely discretionary policy and adhering to strict targets?
17. Under what circumstances would adherence to inflation targets lead to (a) more stable interest rates, (b) less stable interest rates than pursuing discretionary demand management policy?
18. Why might market-orientated supply-side policies have undesirable side effects on aggregate demand?
19. What types of tax cut are likely to create the greatest (a) incentives, (b) disincentives to effort?
20. In what ways can interventionist supply-side policy work *with* the market, rather than against it? What are the arguments for and against such policy?

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

13.1 **Banks, taxes and the fiscal costs of the financial crisis.** A discussion of the government's financial interventions during the financial crisis and their impact on the public finances.

13.2 **The national debt.** This explores the question of whether it matters if a country has a high national debt.

13.3 **Fine-tuning in 1959 and 1960.** This looks at two Budgets in the era of Keynesian 'fine-tuning'.

13.4 **Trends in public expenditure.** This case examines patterns in public expenditure in the UK.

13.5 **The crowding-out effect.** The circumstances in which an increase in public expenditure can replace private expenditure.

13.6 **Injections against the contagion.** The use of discretionary fiscal policy in the late 1990s.

13.7 **Any more G and T?** Did the Code for Fiscal Stability mean that the UK government balanced its books? An examination of the evidence.

13.8 **Assessing PFI.** Has this been the perfect solution to funding investment for the public sector without raising taxes?

13.9 **Discretionary fiscal policy in Japan.** Attempts by successive Japanese governments since 1992 to bring the economy out of recession through expansionary fiscal policy.

13.10 **From the corridor to the floor.** An examination of the Bank of England's monetary policy framework before and after the global financial crisis.

13.11 **The Bank of England's response to the financial crisis.** Focusing on the Bank of England, this case shows the timeline of the responses from 2007.

13.12 **Central banking and monetary policy in the USA and its response to the credit crunch.** This case examines how the Fed conducts monetary policy.

13.13 **Goodhart's Law.** An examination of the difficulty of controlling aggregate demand by setting targets for the money supply.

13.14 **Credit and the business cycle.** This case traces cycles in the growth of credit and relates them to the business cycle. It also looks at some of the implications of the growth in credit.

13.15 **Effective monetary policy versus banking efficiency and stability.** This case examines potential conflicts between banking stability, efficiency and the effective operation of monetary policy.

13.16 **Should central banks be independent of government?** An examination of the arguments for and against independent central banks.

13.17 **Managing the macroeconomy.** This considers whether there have been conflicts of objectives in recent UK macroeconomic policy.

13.18 **Interest rate responses and the financial crisis of 2007/8.** A comparison of the policy responses of the Fed, the ECB and the Bank of England to the credit crunch.

13.19 **Monetary targeting: its use around the world.** This examines the types of monetary targets used around the world.

13.20 **Fiscal and monetary policy in the UK.** An historical overview of UK fiscal and monetary policy.

13.21 **The supply-side revolution in the USA.** 'Reaganomics' and the birth of radical-right supply-side policy in the USA.

13.22 **Alternative approaches to training and education.** This compares the approaches to training and education – a crucial element in supply-side policy – in the UK, France, Germany and the USA.

13.23 **Assistance to small firms in the UK.** An examination of UK policies to encourage investment and growth in the small- and medium-sized-firm sector.

13.24 **Small-firm policy in the EU.** This looks at the range of support available to small and medium-sized firms in the EU.

13.25 **Welfare to work.** An examination of the policy of the UK Labour government (1997–2010) whereby welfare payments were designed to encourage people into employment.

WEB APPENDIX

13.1 ***ISLM* analysis of fiscal and monetary policy.** This appendix uses the *ISLM* model to show the effects of fiscal and monetary policy on national income according to various assumptions about the *IS* and *LM* curves. (For an explanation of *ISLM* analysis, see Web Appendix 11.2.)

WEBSITES RELEVANT TO PART C

Numbers and sections refer to websites listed in the Websites Appendix at the end of the text and hotlinked from this book's website at www.pearsoned.co.uk/sloman.

- For news articles relevant to Part C, search for the *Sloman Economics News* site or follow the News Items link from the student website or MyLab Economics.
- For general news on macroeconomic issues and policy see websites in section A, and particularly A1–5, 7–13, 21, 35, 36. See also links to newspapers worldwide in A38, 39, 43 and 44 and the news search feature in Google at A41. See also links to economics news in A42.
- For general news on money, banking and finance, see websites in section A, and particularly A20–23, 25, 26, 36. See also links to economic and financial news in A42.
- For data on economic growth, employment/unemployment, inflation and the business cycle, see links in B1; also see B4, 12 and 35. For UK data, see B2,3, 5 and 34. For EU data, see B38, 49 and 50. For US data, see B17 and 25. For international data, see B15, 21, 24, 31, 35, 43; H3. For links to data sets, see B1,28, 46; I14.
- For national income statistics for the UK, see B3: search in *Economy* or use search box.
- For the Human Development Index, see site H17.
- For monetary and financial data (including data for money supply and interest rates), see section F and particularly F2. Note that you can link to central banks worldwide from site F17. See also the links in B1.
- For information on the development of ideas, see C12, 18.
- For a model of the economy (based on the Treasury model), see *The Virtual Chancellor* (site D1). The model allows you devise and test your own Budget.
- For information on UK fiscal policy and government borrowing, see sites E30, 36; F2. See also sites A1–8 at Budget time. For fiscal policy in the eurozone, see G13 and G1.

(Continued)

- For monetary policy in the UK, see F1 and E30. For monetary policy in the eurozone, see F6 and 5. For monetary policy in the USA, see F8. For monetary policy in other countries, see the respective central bank site in section F and the links in F17.
- For demand-side policy in the UK, see the latest Budget Report (e.g. section on maintaining macroeconomic stability) at site E30.
- For inflation targeting in the UK and eurozone see sites F1 and 6.
- For the current approach to UK supply-side policy, see the latest Budget Report (e.g. sections on productivity and training) at site E30. See also sites E5 and 9. For European policy see sites G5, 7, 9, 12, 14, 19.
- For support for a market-orientated approach to supply-side policy see C17 and E34.
- For information on training in the UK and Europe, see sites E5, 10; G5, 14.
- For information on the support for small business in the UK see site E38.
- For student resources relevant to this Part, see sites C1–7, 9, 10, 12, 19.

International economics

Chapter 14

Globalisation and international trade

Countries operate in a global economy. In these final two chapters we explore global economic relationships. In particular, we examine the importance of international trade (this chapter) and finance (next chapter) for national economies. We begin this chapter by examining the nature of the interdependencies between countries in the flows of both goods and finance.

We then focus on trade. Trade between nations has the potential to benefit all participating countries (albeit to differing extents). Many people, however, argue strongly for restrictions on trade, seeing unrestrained imports as a threat. But are they justified in fearing international competition, or are they merely trying to protect some vested interest at the expense of everyone else? Section 14.3 examines the arguments for restricting trade.

If there are conflicting views as to whether we should have more or less trade, what has been happening on the world stage? Section 14.4 looks at the various moves towards making trade freer and at the obstacles that have been met.

A step on the road to freer trade is for countries to enter free-trade agreements with just a limited number of other countries. An example is the United States–Mexico–Canada agreement (USMCA). We consider such 'preferential trading systems' in Section 14.5. Then, in Section 14.6, we look at probably the world's most famous preferential trading system, the European Union, and, in particular, at the development of a 'single European market'.

In Section 14.7 we consider some of the possible economic implications for the UK following its departure from the EU, culminating with the EU–UK Trade and Cooperation Agreement, which took effect from 1 January 2021. The focus here is on the supply-side effects of the UK's new trading relationship with the EU and the rest of the world.

Finally we examine the role of trade for developing countries. Does trade with the rich world help them to develop, or does it merely result in their domination by rich countries and giant multinational companies?

After studying this chapter, you should be able to answer the following questions:

- How can economic and financial interdependencies create international business cycles?
- How has international trade grown over the years? Have countries become more or less interdependent?
- What are the benefits from international trade?
- Which goods should a country export and which should it import?
- Why do countries often impose restrictions on trade? Why does the World Trade Organization (WTO) try to get trade restrictions reduced or even eliminated?
- Why do countries form free-trade areas and other types of trading alliance?
- How has the 'single market' in the EU benefited its members?
- What supply-side effects could there be on the UK economy following its departure from the EU?
- What approaches should developing countries adopt towards trade with rich countries?

14.1 GLOBAL INTERDEPENDENCE

How countries are connected through trade and finance

We live in an interdependent world. Countries are affected by the economic health of other countries and by their governments' policies. Problems in one part of the world can spread like a contagion to other parts, with perhaps no country immune. This was clearly illustrated by the credit crunch of 2007/8. A crisis that started in the sub-prime market in the USA soon snowballed into a worldwide recession.

There are two major ways in which this process of 'globalisation' affects individual economies. The first is through trade. The second is through financial markets.

Interdependence through trade

So long as nations trade with one another, the domestic economic actions of one nation will have implications for those which trade with it. For example, if the US administration feels that the US economy is growing too fast, it might adopt various contractionary fiscal and monetary measures, such as higher tax rates or interest rates. US consumers will not only consume fewer domestically produced goods, but also reduce their consumption of imported products. But US imports are other countries' exports. A fall in these other countries' exports will lead to a multiplier effect in these countries. Output and employment will fall.

Changes in aggregate demand in one country thus send ripples throughout the global economy. The process whereby changes in imports into (or exports from) one country affect national income in other countries is known as the **international trade multiplier**.

TC 14 p 249

Pause for thought

Assume that the US economy expands. What will determine the size of the multiplier effect on other countries?

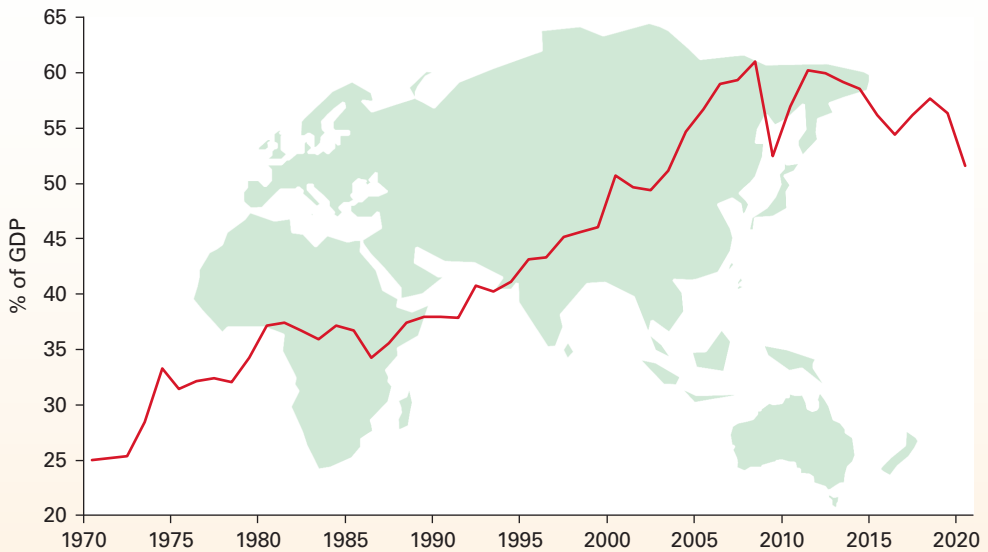
The more open an economy, the more vulnerable it will be to changes in the level of economic activity in the rest of the world. This problem will be particularly acute if a nation is heavily dependent on trade with one other nation (e.g. Canada on the USA) or one other region (e.g. Switzerland on the EU).

Until very recently, international trade had been growing as a proportion of countries' national income for many years. This is illustrated in Figure 14.1, which shows the ratio of the sum of global exports

Definition

International trade multiplier The effect on national income in country B of a change in exports (or imports) of country A.

Figure 14.1 World trade in goods and services (imports plus exports) as a % of GDP



Source: series NE.TRD.GNFS.ZS (World Bank)

BOX 14.1 TRADE IMBALANCE IN THE USA AND CHINA

An illustration of economic and financial interdependencies

In 2021, the USA had a current account deficit of over $800 billion – the equivalent of 3.5 per cent of GDP (see chart). American current account deficits are not new. Since 1997, the US current account deficit has averaged 3.3 per cent of GDP, reaching as high as 5.9 per cent in 2006.

The current account deficit is offset by an equal and opposite capital-plus-financial account surplus, much of which consists of the purchase of US government bonds and Treasury bills. These massive inflows to the USA are thought to represent some three-quarters of all the savings which the rest of the world invests abroad. These financial inflows have permitted the persistence of sizeable US current account deficits.

To attract such large inflows, it might be expected that US interest rates would have to be high. Yet US interest rates were consistently low from the late 1990s and, of course, exceptionally so during the global financial crisis and the COVID-19 pandemic. How is it, then, that with such low interest rates, the USA managed to attract such vast inflows of finance and thereby maintain such a large financial account surplus?

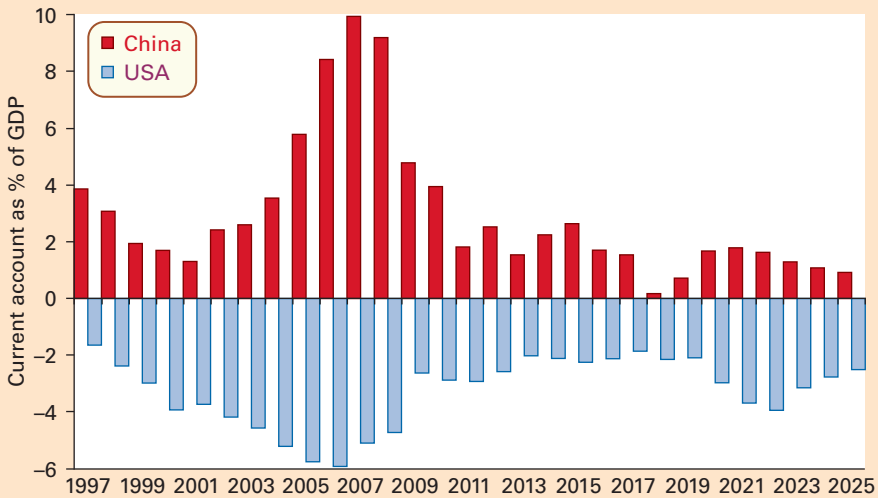
China's appetite for dollars

Several Asian countries, including China, pegged their currencies to the dollar and had been running large current account surpluses. For instance, the Chinese current account surplus was typically around 5 to 10 per cent of GDP during the mid- to late 2000s (see chart). Instead of letting the Chinese yuan (or 'renminbi') appreciate against the dollar, the Chinese central bank used the surpluses to buy dollars.

There were three perceived advantages in doing this.

- First, it allowed China to *build up reserves* and thereby bolster its ability to resist any future speculative attacks

US and Chinese current account (% of GDP)



Note: Figures from 2022 figures are forecasts

Source: Based on data from *World Economic Outlook* database (IMF, October 2022)

and imports of goods and services to global GDP since 1970. Over the period, the ratio has risen from around 25 per cent to around 60 per cent.

Most nations have been committed to freer trade. This has been aided by the World Trade Organization (see Section 14.4) overseeing the dismantling of trade barriers. However, the increased economic and financial interdependence made countries more vulnerable to world trade fluctuations and disruptions, such as the global financial crisis and the COVID-19 pandemic. World output fell by 2 per cent and trade volumes by 10 per cent in 2009; in 2020 output fell by 3 per cent and trade volumes by 8 per cent.

Consequently, doubts have risen over the benefits of freer trade. With a growth in protectionist rhetoric (e.g. under the Trump administration) and concerns about the vulnerability of countries to disruptions in trade, such as those arising from the COVID-19

Pause for thought

Are exports likely to resume growing faster than GDP? What will determine the outcome?

CASE STUDIES & APPLICATIONS

on its currency. Chinese foreign exchange reserves rose from around $170 billion in 2000 to $4 trillion in 2014 – a staggering 24-fold increase. The effect was a huge increase in global liquidity and hence money supply.

- Second, and more important, it *kept their exchange rates low* and thereby helped to keep their exports competitive. This helped to sustain their rapid rates of economic growth.
- Third, it helped to *keep US interest rates down* and therefore boost US spending on Asian exports.

In 2005, the Chinese, after much international pressure, agreed to revalue the yuan and would then peg it against a basket of currencies with subsequent further revaluations. Between July 2005 and July 2008 the yuan was allowed to appreciate by around 18 per cent against the US dollar while the exchange rate index rose by 7 per cent (the exchange rate index of the US dollar fell by 15 per cent).

But, with the global economic downturn biting in 2008 and concerns about slowing Chinese export growth, the Chinese authorities effectively fixed the yuan once again.

This remained the case until June 2010 when again the yuan was revalued. Between June 2010 and January 2014 the yuan appreciated a further 12 per cent against the US dollar. Therefore, over the period from July 2005 to January 2014 the yuan had appreciated by around 33 per cent against the US dollar.

China's economic slowdown

The managed appreciation of the yuan, coupled with the slowdown of the global economy, saw China's current account surplus begin to wane, falling from 9 per cent of GDP in 2008 to $1\frac{1}{2}$ per cent by 2013. Meanwhile, though still sizeable, the USA's current account deficit had shrunk. By 2013 the deficit had fallen to 2.1 per cent of GDP – less than half the level of the mid 2000s.

China's economic growth rate, which had been as high as 14 per cent in 2007, fell back to between 6 and 8 per cent from 2012. This weakening of growth reflected a sharp decline in the growth of China's exports of goods and services. While export volumes grew by 26 per cent per annum from 2002 to 2007, export growth fell to single figures from 2012, with a contraction in 2015.

In late 2014 China's central bank began cutting interest rates in response to weaker growth. Then, in August 2015, it took the step to devalue the yuan by 2 per cent against the US dollar. However, this raised concerns of further devaluations and loosening of monetary policy. Hence, the yuan began falling more sharply, forcing the central bank to sell dollars to halt the depreciation.

The yuan began to fall again in 2018 as export growth weakened and the Trump administration began imposing tariffs on a range of Chinese imports in response to its persistent trade deficit with China. As China responded with tariffs of its own, fears grew of a 'tit-for-tat' trade war. Then, in August 2019, as the yuan rose above the important threshold of 7 to the dollar (making Chinese exports cheaper in the USA), the USA declared China a 'currency manipulator'.

Tensions eased, at least temporarily, when the USA and China signed a 'Phase 1' trade deal in 2020. This included a commitment from China not to engage in competitive devaluations and commitments to increase purchases of US farm and manufactured goods, energy and services. The latter proved difficult to meet in the face of the pandemic and subsequent supply-chain issues. Then, with rising global political tensions, the prospects for further substantial progress on trade co-operation appeared to be on hold.

TC 11 p 61

Examine the merits for the Chinese of (a) floating the yuan freely; (b) pegging it to a trade-weighted basket of currencies.

Download from the IMF World Economic Outlook Database the annual percentage change in GDP at constant prices for China, the UK, the USA and Germany. Construct a line chart showing economic growth in these countries across time and then prepare a short note describing what your chart shows.

pandemic and the Russian invasion of Ukraine, world trade may have peaked as a proportion of GDP – at least for the time being.

Financial interdependence

International trade has grown rapidly, but international financial flows have grown much more rapidly. Trillions of dollars are traded across the foreign exchanges each day. Many of the transactions are short-term financial flows, moving to where interest rates are most favourable or to currencies where the exchange rate is likely to appreciate. This again makes countries interdependent.

TC 3 p 10

Financial interdependency impacts not only on financial institutions in different countries, but also on national economies. It also illustrates how global responses can be needed. As a result of the credit crunch world leaders were seriously worried that the whole world would plunge into deep recession. A co-ordinated policy response from governments and central banks began in earnest in October 2008 when governments in Britain, Europe, North America and other parts of the world injected some $2 trillion of extra capital into banks.

International business cycles

There is an old saying: 'If America sneezes, the rest of the world catches a cold.' Viruses of a similar nature regularly infect the world economy. The credit crunch of the 2007/8 resulting from defaults on US sub-prime debt (see Box 11.2) was a dramatic example of this. As a consequence of both trade and financial interdependence, the world economy, like the economy of any individual country, tends to experience periodic fluctuations in economic activity – an *international* business cycle. The implication of this is that countries will tend to share common problems and concerns at the same time. At one time, the most pressing problem may be world inflationary pressures; at another time, it may be a world recession.

TC 14 p 249

In order to avoid 'beggar-my-neighbour' policies, it is better to seek *common* solutions to these common problems: i.e. solutions that are international in scope and design rather than narrowly based on national self-interest. For example, during a world recession, countries are likely to suffer from rising unemployment. Policies that lead to a depreciation of the exchange rate (such as cutting interest rates) will help to stimulate demand by making exports cheaper and imports more expensive. But this will then only worsen the trade balance of other countries, whose aggregate demand will thus fall. The first country is thus tackling its own unemployment at the expense of rising unemployment in other countries.

However, if other nations (which will also be experiencing higher unemployment) can be convinced to co-ordinate their policy actions, an expansionary *international* economic policy will benefit all. In addition to the resulting rise in their imports, all nations will also experience rising export sales.

Even if national policies are not in the strictest sense co-ordinated, discussions between nations regarding the nature and magnitude of the problems they face may help to improve the policy-making process.

Global policy response

Global economic interdependence and an international business cycle can aid the process of international co-operation between countries. Countries frequently meet in various groupings – from the narrow group of the world's seven richest developed countries (the G7, which includes the USA, Japan, Germany, the UK, France, Italy and Canada) to broader groups such as the G20, which, in addition to the G7 and other rich countries also includes larger developing countries, such as China, India, Brazil and South Africa.

Global interdependence raises important questions about the role of international organisations like the World Trade Organization (WTO) and the International Monetary Fund (IMF).

The World Trade Organization's role is to encourage freer trade. We consider the WTO in Section 14.4.

The IMF's remit is to promote global growth and stability, to help countries through economic difficulty and to help developing economies achieve macroeconomic stability and reduce poverty. In response to the global economic and financial crisis of the late 2000s, the IMF's budget was substantially increased. Then, in 2021, during the COVID-19 pandemic, member countries and central banks agreed to double the amount they stood ready to lend to the IMF.

Recap

1. Changes in aggregate demand in one country will affect the amount of imports purchased and thus the amount of exports sold by other countries and hence their national income. There is thus an international trade multiplier effect.
2. The world is also financially interdependent, with huge flows of finance flowing from one country to another. This makes all countries susceptible to financial shocks, such as the credit crunch of 2007/8.
3. To prevent problems in one country spilling over to other countries and to stabilise the international business cycle will require co-ordinated policies between nations and the intervention of various international agencies, such as the IMF.

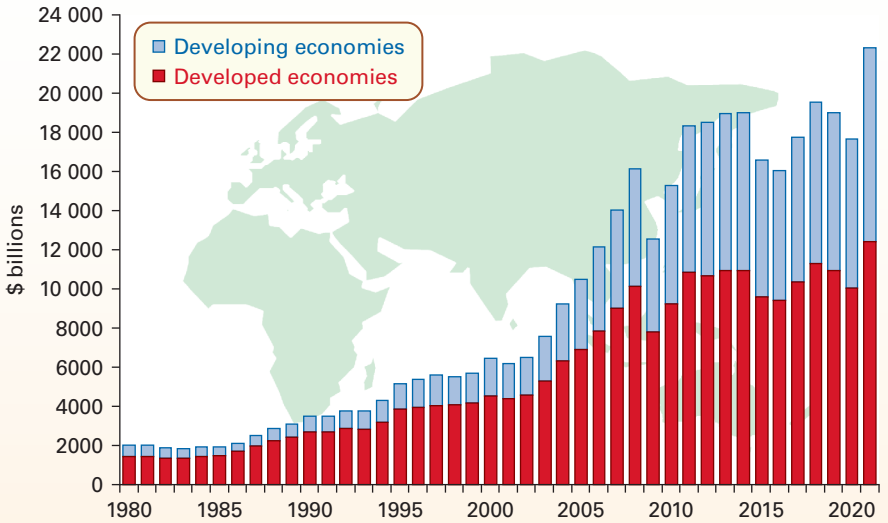
14.2 THE ADVANTAGES OF TRADE

Can international trade make all countries better off?

Trading patterns

Developed economies have, until recently, dominated world trade. Figure 14.2 shows the value of world exports since 1980 originating from developed countries and developing countries (including the former Soviet republics). In the 1980s and 1990s the percentage of world merchandise exports originating from developed countries averaged 75 per cent; in the 10-year period up to 2021 it averaged 58 per cent.

Figure 14.2 World merchandise exports by value, $ billions



Note: Developing countries include Commonwealth of Independent States (former Soviet republics) from 1992

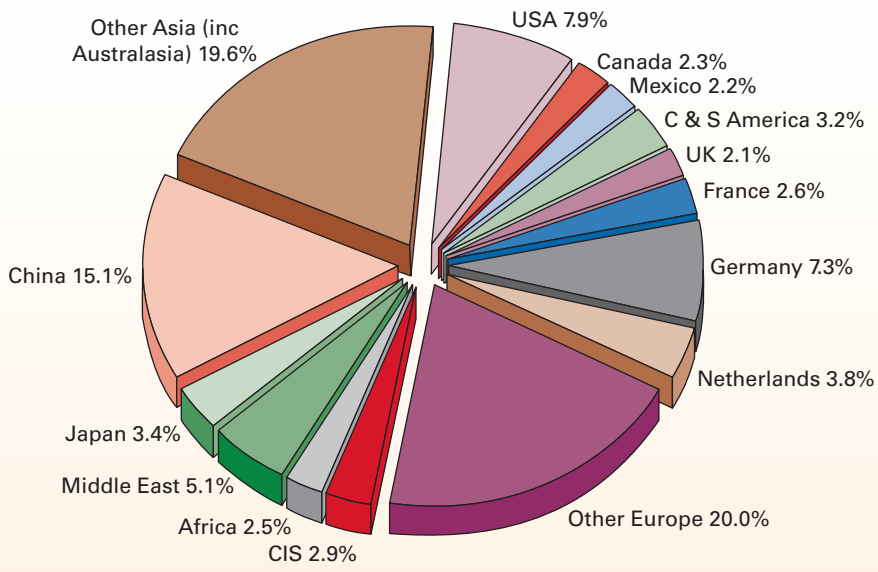
Source: Based on data from UNCTADStat (UNCTAD, 2022)

The reason for this changing global pattern of trade is that many of the countries with the most rapid *growth* in exports can now be found in the developing world. The growth in merchandise exports from the group of developing nations collectively known as the BRICS[1] (Brazil, Russia, India, China and South Africa) has been especially rapid. Between them they accounted for just 5.5 per cent of the value of world exports in 1992. In 2021 they accounted for 20.9 per cent of world exports. More recently, other countries, such as Mexico, Turkey, Indonesia, Cambodia and Vietnam, have joined the ranks of rapidly growing 'newly industrialised' developing countries.

Figure 14.3 shows the geographical origin of exports in 2021 in clearer focus. Despite more rapid growth in

[1]Sometimes the term is used to refer just to the first four countries. When South Africa is excluded, the term is written BRICs rather than BRICS.

Figure 14.3 Share of world merchandise exports, by value (2021)



Source: Based on data in *WTO Stats* (WTO, 2022)

BOX 14.2 **TRADING PLACES** CASE STUDIES & APPLICATIONS

Partners in trade

(a) Top trading countries by value and world share, 2021

Rank	Exporters	$ billion	Share	Rank	Importers	$ billion	Share
1	China	3364	15.1	1	USA	2937	13.0
2	USA	1755	7.9	2	China	2688	11.9
3	Germany	1632	7.3	3	Germany	1419	6.3
4	Netherlands	836	3.8	4	Japan	769	3.4
5	Japan	756	3.4	5	Netherlands	757	3.4
14	UK	468	2.1	8	UK	694	3.1

Source: Based on data in *WTO Stats* (WTO, 2022)

(b) Merchandise trade by destination and origin, 2020

	China		USA		Germany		Japan		UK	
	Export destinations	**%**	**Export destinations**	**%**	**Export destinations**	**%**	**Export destinations**	**%**	**Export destinations**	**%**
1	USA	17.5	Canada	17.8	EU-26	51.7	China	22.0	EU-27	46.2
2	EU-27	15.1	EU-27	16.2	USA	8.6	USA	18.5	USA	14.4
3	Hong Kong	10.5	Mexico	14.9	China	8.0	EU-27	9.2	Switzerland	5.0
4	Japan	5.5	China	8.7	UK	5.5	S Korea	7.0	China	4.7
5	Vietnam	4.4	Japan	4.5	Switzerland	4.8	Taiwan	6.9	Hong Kong	2.7
	Import origins	**%**	**Import origins**	**%**	**Import origins**	**%**	**Import origins**	**%**	**Import origins**	**%**
1	EU-27	12.6	China	19.0	EU-26	52.0	China	25.8	EU-27	46.3
2	Taiwan	9.8	EU-27	17.6	China	11.4	EU-27	11.4	China	11.9
3	Japan	8.5	Mexico	13.7	USA	6.7	USA	11.3	USA	9.2
4	S Korea	8.4	Canada	11.5	Switzerland	4.6	Australia	5.6	Russia	3.9
5	USA	6.6	Japan	5.1	UK	3.3	Taiwan	4.1	Canada	2.4

Source: *Trade Profiles, Statistics database* (WTO, 2022)

Table (a) shows the world's top merchandise exporters and importers by value in 2021. The UK was the 14th largest exporter and the eighth largest importer by value.

But who do countries trade with? Table (b) looks at the trading partners of four of the largest trading nations (by exports) and also those of the UK. The table shows that trading blocs (see Section 14.5) and geography matter. For instance, in the case of Germany over one-half of its trade is with other EU countries. Similarly, one-third of the value of the USA's exports and one-quarter of its imports can be attributed to Canada and Mexico, fellow members of the United States–Mexico–Canada Agreement (USMCA), the successor to the North American Free Trade Agreement (NAFTA) (see pages 433–4).

1. *What factors might be expected to increase the relative size of trade flows between any two countries?*
2. *Does the fact that world trade has increased at a much faster rate than world GDP highlight the limitations of trade as a driver of economic growth?*

Download the trade profiles of the countries in the table (b) from Statistics database of the WTO. Compile a short briefing paper detailing the exports and imports of each country by the types of commodities.

trade values in other regions, Europe remains an important geographical centre for trade. In 2021 it accounted for 35.8 per cent of world exports by value (and was the destination for 35.5 per cent of imports). While Africa as a whole has experienced significant growth in trade since the early 1990s, many of the poorest African countries have seen negligible growth in trade over the period. Hence, in 2021 Africa accounted for only 2.5 per cent of world exports by value (and was the destination for 2.8 per cent of imports).

Box 14.2 considers in more detail some patterns and trends in world trade.

We now go on to consider the reasons why countries trade with each other.

Specialisation as the basis for trade

KI 1 p 5

What motivates countries to trade with each other? What do they gain out of it? The reasons for international trade are really only an extension of the reasons for trade *within* a nation. Rather than people trying to be self-sufficient and do everything for themselves, it makes sense to specialise.

Firms specialise in producing certain types of goods. This allows them to gain economies of scale and to exploit their entrepreneurial and management skills and the skills of their labour force. It also allows them to benefit from their particular location and from the ownership of any particular capital equipment or other assets they might possess. With the revenues that firms earn, they buy in the inputs that they need from other firms and the labour they require. Firms thus trade with each other.

TC 9 p 45

Countries also specialise. They produce more than they need of certain goods. What is not consumed domestically is exported. The revenues earned from the exports are used to import goods that are not produced in sufficient amounts at home.

But which goods should a country specialise in? What should it export and what should it import? The answer is that it should specialise in those goods in which it has a *comparative advantage*. Let us examine what this means.

The law of comparative advantage

Countries have different endowments of factors of production. They differ in population density, labour skills, climate, raw materials, capital equipment, etc. These differences tend to persist because factors are relatively immobile between countries. Obviously land and climate are totally immobile, but even with labour and capital there tend to be more restrictions (physical, social, cultural or legal) on their international movement than on their movement within countries. Thus the ability to supply goods differs between countries.

What this means is that the relative costs of producing goods will vary from country to country. For example, one country may be able to produce 1 fridge for the same cost as 6 tonnes of wheat or 3 games consoles, whereas another country may be able to produce 1 fridge for the same cost as only 3 tonnes of wheat but 4 games consoles. It is these differences in relative costs that form the basis of trade.

At this stage we need to distinguish between *absolute advantage* and *comparative advantage*.

Absolute advantage

When one country can produce a good with fewer resources than another country it is said to have an **absolute advantage** in that good. If France can produce grapes with fewer resources than the UK, and the UK can produce barley with fewer resources than France, then France has an absolute advantage in grapes and the UK an absolute advantage in barley. Production of both grapes and barley will be maximised by each country specialising and then trading with the other country. Both will gain.

Comparative advantage

The above seems intuitive, but trade between two countries can still be beneficial even if one country could produce all goods with fewer resources than the other, providing the *relative* efficiency with which goods can be produced differs between the two countries.

Take the case of a developed country that is absolutely more efficient than a less developed country at producing both wheat and cloth. Assume that with a given amount of resources (labour, land and capital) the alternatives shown in Table 14.1 can be produced in each country

Despite the developed country having an absolute advantage in both wheat and cloth, the less developed

Definition

Absolute advantage A country has an absolute advantage over another in the production of a good if it can produce it with less resources than the other country can.

Table 14.1 Production possibilities for two countries

	Kilos of wheat		Metres of cloth
Less developed country	Either 2	or	1
Developed country	Either 4	or	8

country (LDC) has a **comparative advantage** in wheat, and the developed country has a *comparative* advantage in cloth. This is because wheat is relatively cheaper in terms of cloth in the LDC: only 1 metre of cloth has to be sacrificed to produce 2 kilos of wheat, whereas 8 metres of cloth would have to be sacrificed in the developed country to produce 4 kilos of wheat. In other words, the opportunity cost of wheat is 4 times higher in the developed country (8/4 compared with 1/2).

Pause for thought

Draw up a table similar to Table 14.1, only this time assume that the figures are: LDC 6 wheat or 2 cloth; DC 8 wheat or 20 cloth. What are the opportunity cost ratios now?

On the other hand, cloth is relatively cheaper in the developed country. Here the opportunity cost of producing 8 metres of cloth is only 4 kilos of wheat, whereas in the LDC 1 metre of cloth costs 2 kilos of wheat. Thus the opportunity cost of cloth is 4 times higher in the LDC (2/1 compared with 4/8).

TC 1 p 7

To summarise: countries have a comparative advantage in those goods that can be produced at a lower opportunity cost than in other countries.

If countries are to gain from trade, they should export those goods in which they have a comparative advantage and import those goods in which they have a comparative disadvantage. Given this we can state a **law of comparative advantage.**

KEY IDEA 36

The law of comparative advantage. **Provided opportunity costs of various goods differ in two countries, both of them can gain from mutual trade if they specialise in producing (and exporting) those goods that have relatively low opportunity costs compared with the other country's.**

But why do they gain if they specialise according to this law? And just what will that gain be? We consider these questions next.

The gains from trade based on comparative advantage

Before trade, unless markets are very imperfect, the prices of the two goods are likely to reflect their opportunity costs. For example, in Table 14.1, since the less developed country can produce 2 kilos of wheat for 1 metre of cloth, the *price* of 2 kilos of wheat will roughly equal 1 metre of cloth.

TC 1 p 7

Assume, then, that the pre-trade exchange ratios of wheat for cloth are as follows:

LDC:	2 wheat for 1 cloth
Developed country:	1 wheat for 2 cloth (i.e. 4 for 8)

Both countries will now gain from trade, provided the exchange ratio is somewhere between 2:1 and 1:2. Assume, for the sake of argument, that it is 1:1: that 1 wheat trades internationally for 1 cloth. How will each country gain?

TC 2 p 9

The LDC gains by exporting wheat and importing cloth. At an exchange ratio of 1:1, it now has to give up only 1 kilo of wheat to obtain 1 metre of cloth, whereas before trade it had to give up 2 kilos of wheat.

The developed country gains by exporting cloth and importing wheat. Again at an exchange ratio of 1:1, it now has to give up only 1 metre of cloth to obtain a kilo of wheat, whereas before it had to give up 2 metres of cloth.

Thus both countries have gained from trade.

TC 9 p 45

The actual exchange ratios will depend on the relative prices of wheat and cloth after trade takes place. These prices will depend on total demand for and supply of the two goods. It may be that the trade exchange ratio is nearer to the pre-trade exchange ratio of one country than the other. Thus the gains to the two countries need not be equal.

The limits to specialisation and trade

Does the law of comparative advantage suggest that countries will completely specialise in just a few products? In practice, countries are likely to experience *increasing* opportunity costs. The reason for this is that, as a country increasingly specialises in one good, it will have to use resources that are less and less suited to its production and which were more suited to other goods. Thus ever-increasing amounts of the other goods will have to be sacrificed. For example, as a country specialises more and more in grain production, it will have to use land that is less and less suited to growing grain.

TC 1 p 7

These increasing costs as a country becomes more and more specialised will lead to the disappearance of its comparative cost advantage. When this happens, there will be no point in further specialisation. Thus whereas a country like Germany has a comparative advantage in capital-intensive manufactures, it does not produce only manufactures. It would make no sense not to use its fertile lands to produce food or its forests to produce timber. The opportunity costs of diverting all agricultural labour to industry would be very high.

Definitions

Comparative advantage A country has a comparative advantage over another in the production of a good if it can produce it at a lower opportunity cost: i.e. if it has to forgo less of other goods in order to produce it.

Law of comparative advantage Trade can benefit all countries if they specialise in the goods in which they have a comparative advantage.

Other reasons for gains from trade

Decreasing costs. Even if there are no initial comparative cost differences between two countries, it will still benefit both to specialise in industries where economies of scale can be gained, and then to trade. Once the economies of scale begin to appear, comparative cost differences will also appear, and thus the countries will have gained a comparative advantage in these industries.

This reason for trade is particularly relevant for small countries where the domestic market is not large enough to support large-scale industries. Thus exports form a much higher percentage of GDP in small countries such as Singapore than in large countries such as the USA.

Differences in demand. Even with no comparative cost differences and no potential economies of scale, trade can benefit both countries if demand conditions differ.

If people in country A like beef more than lamb, and people in country B like lamb more than beef, then rather than A using resources better suited for lamb to produce beef, and B using resources better suited for producing beef to produce lamb, it will benefit both to produce beef *and* lamb and to export the one they like less in return for the one they like more.

Increased competition. If a country trades, the competition from imports may stimulate greater efficiency at home. This extra competition may prevent domestic monopolies/oligopolies from charging high prices. It may stimulate greater research and development and the more rapid adoption of new technology. It may lead to a greater variety of products being made available to consumers.

Trade as an 'engine of growth'. In a growing world economy, the demand for a country's exports is likely to grow over time, especially when these exports have a high income elasticity of demand. This will provide a stimulus to growth in the exporting country.

Non-economic advantages. There may be political, social and cultural advantages to be gained by fostering trading links between countries.

The terms of trade

What price will our exports fetch abroad? What will we have to pay for imports? The answer to these questions is given by the terms of trade. The **terms of trade** are defined as:

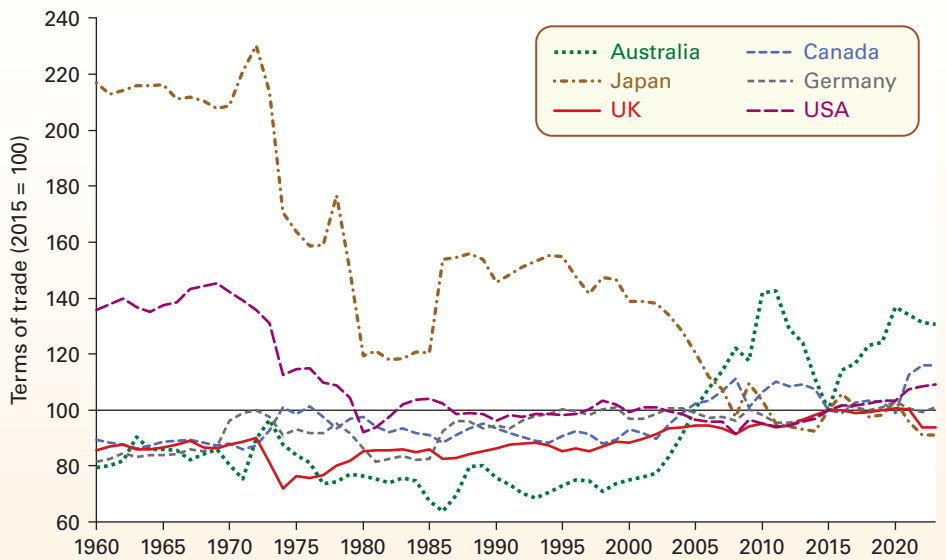
$$\frac{\text{Average price of exports}}{\text{Average price of imports}}$$

expressed as an index, where prices are measured against a base year in which the terms of trade are assumed to be 100. Thus, if the average price of exports relative to the average price of imports has risen by 25 per cent since the base year, the terms of trade will now be 125. The terms of trade for selected countries are shown in Figure 14.4 (with 2015 as the base year).

Definition

Terms of trade The price index of exports divided by the price index of imports and then expressed as a percentage. This means that the terms of trade will be 100 in the base year.

Figure 14.4 Terms of trade for goods and services (2015 = 100)



Note: Data from 2022 are based on forecasts

Source: Based on data in *AMECO Database* (European Commission, DGECFIN, May 2022)

If the terms of trade rise (export prices rising relative to import prices), they are said to have 'improved', since fewer exports now have to be sold to purchase any given quantity of imports. Changes in the terms of trade are caused by changes in the demand and supply of imports and exports and by changes in the exchange rate.

Pause for thought

In Figure 14.4, which countries have seen an improvement in their terms of trade in recent years?

Recap

1. Economies have become increasingly interdependent as a result of the increase in world trade and the interdependencies of countries' financial systems.
2. Countries can gain from trade if they specialise in producing those goods in which they have a comparative advantage: i.e. those goods that can be produced at relatively low opportunity costs. This is merely an extension of the argument that gains can be made from the specialisation and division of labour.
3. If two countries trade, then, provided that the trade price ratio of exports and imports is somewhere between the pre-trade price ratios of these goods in the two countries, both countries can gain.
4. With increasing opportunity costs there will be a limit to specialisation and trade. As a country increasingly specialises, its (marginal) comparative advantage will eventually disappear.
5. The terms of trade give the price of exports relative to the price of imports expressed as an index, where the base year is 100.
6. Gains from trade also arise from decreasing costs (economies of scale), differences in demand between countries, increased competition from trade and the transmission of growth from one country to another. There may also be non-economic advantages from trade.

14.3 ARGUMENTS FOR RESTRICTING TRADE

If trade can benefit everyone, then why do countries attempt to limit trade?

We have seen how trade can bring benefits to all countries. But when we look around the world we often see countries erecting barriers to trade. Their politicians know that trade involves costs as well as benefits.

Possible barriers to imports include the following:

- **tariffs** (i.e. customs duties) on imports;
- *quotas* (i.e. restrictions on the amount of certain goods that can be imported);
- *subsidies on domestic products* to give them a price advantage over imports;
- *administrative regulations* designed to exclude imports, such as customs delays or excessive paper work;
- *procurement procedures* whereby governments favour domestic producers when purchasing equipment (e.g. defence equipment).

Alternatively, governments may favour domestic producers by subsidising their exports in a process known as **dumping**. The goods are 'dumped' at artificially low prices in the foreign market.

In looking at the costs and benefits of trade, the choice is not the stark one of whether to have free trade or no trade at all. Although countries may sometimes contemplate having completely free trade, typically countries limit their trade. However, they certainly do not ban it altogether.

Arguments in favour of restricting trade

Arguments having some general validity

TC 7 p 26

The infant industry argument. Some industries in a country may be in their infancy but have a potential comparative advantage. This is particularly likely in developing countries. Such industries are too small yet to have gained economies of scale; their workers are inexperienced; there is a lack of back-up facilities – communications networks, specialist research and development, specialist suppliers, etc. – and they may have only limited access to finance for expansion. Without protection, these

Definitions

Tariffs Taxes on imported products

Dumping When exports are sold at prices below marginal cost – often as a result of government subsidy.

infant industries will not survive competition from abroad.

Protection from foreign competition, however, will allow them to expand and become more efficient. KI 36 p 422 Once they have achieved a comparative advantage, the protection can then be removed to enable them to compete internationally.

The senile industry argument. This is similar to the infant industry argument. It is where industries with a potential comparative advantage have been allowed to run down and can no longer compete effectively. They may have considerable potential but be simply unable to make enough profit to afford the necessary investment without some temporary protection. This is one of the most powerful arguments used to justify the use of special protection for the automobile and steel industries in the USA.

Pause for thought

How would you set about judging whether an industry had a genuine case for infant/senile industry protection?

To reduce reliance on goods with little dynamic potential. Many developing countries have traditionally exported primaries: foodstuffs and raw materials. The world demand for some of these, however, is fairly income inelastic, and thus tends to grow relatively slowly (see Section 14.7). In such cases, free trade is not an engine of growth. Instead, if it encourages countries' economies to become locked into a pattern of primary production, it may prevent them from expanding in sectors like manufacturing which have a higher income elasticity of demand. There may thus be a valid argument for protecting or promoting manufacturing industry.

To prevent 'dumping' and other unfair trade practices. A country may engage in dumping by subsidising its exports. The result is that prices may no longer reflect comparative costs. Thus the world would benefit from tariffs being imposed by importers to counteract such practice.

It can also be argued that there is a case for retaliating against countries which impose restrictions on your exports. In the *short* run, both countries are likely to be made worse off by a contraction in trade. But if the retaliation persuades the other country to remove its restrictions, it may have a longer-term benefit. In some cases, the mere threat of retaliation may be enough to get another country to remove its protection. TC 7 p 7

KI 19 p 122 *To prevent the establishment of a foreign-based monopoly.* Competition from abroad could drive domestic producers out of business. The foreign company, now having a monopoly of the market, could charge high prices with a resulting misallocation of resources. The problem could be tackled either by restricting imports or by subsidising the domestic producer(s).

All of the above arguments suggest that governments should adopt a 'strategic' approach to trade. **Strategic trade theory** argues that protecting certain industries allows a net gain in the *long* run from increased competition in the market (see Case Study 14.3 on the student website). This argument has been used to justify the huge financial support given to the aircraft manufacturer Airbus, a consortium based in four European countries. The subsidies have allowed it to compete with Boeing, which would otherwise have a monopoly in many types of passenger aircraft. Airlines and their passengers worldwide, it is argued, have benefited from the increased competition.

To spread the risks of fluctuating markets. A highly specialised economy – Zambia with copper, Cuba with sugar – is highly susceptible to world market fluctuations. Greater diversity and greater self-sufficiency can reduce these risks.

To reduce the influence of trade on consumer tastes. It is a mistake to assume that fixed consumer tastes dictate the pattern of production through trade. Multinational companies through their advertising and other forms of sales promotion may influence consumer tastes. Thus some restriction on trade may be justified in order to reduce this 'producer sovereignty'.

To prevent the importation of harmful goods. A country may want to ban or severely curtail the importation of things such as drugs, pornographic literature and live animals.

To take account of externalities. Free trade will tend to reflect private costs. Both imports and exports, however, can involve externalities. The mining of many minerals for export may adversely affect the health of miners; the production of chemicals for export may KI 28 p 196

Definitions

Infant industry An industry that has a potential comparative advantage, but which is as yet too underdeveloped to be able to realise this potential.

Strategic trade theory The theory that protecting/supporting certain industries can enable them to compete more effectively with large monopolistic rivals abroad. The effect of the protection is to increase long-run competition and may enable the protected firms to exploit a comparative advantage that they could not have done otherwise.

involve pollution; the importation of juggernaut lorries may lead to structural damage to houses; shipping involves large amounts of CO_2 emissions (some 3 to 5 per cent of total world emissions).

In recent years some politicians and green groups have called for the imposition of 'carbon tariffs'. The rate of tariff would reflect the amount of carbon emitted in the production of the good being imported. Such tariffs would be hard to implement, however. Assessing and valuing the carbon emitted would be very difficult and could lead to arbitrary tariff rates. Also, domestic goods would have to be subject to similar taxes.

Economic arguments having some validity for specific groups or countries

The arguments considered so far are of general validity: restricting trade for such reasons could be of net benefit to the world. There are two other arguments, however, that are used by individual governments for restricting trade, where their country will gain but at the *expense* of other countries, such that there will be a net loss to the world.

KI 19 p 122

Exploitation of market power. If a country, or a group of countries, has monopsony power in the purchase of imports (i.e. they are individually or collectively a very large economy, such as the USA or the EU), then they could gain by restricting imports so as to drive down their price. Similarly, if countries have monopoly power in the sale of some export (e.g. OPEC countries with oil – see Box 6.3 on page 142), then they could gain by restricting exports, thereby forcing up the price (see Web Appendix 14.2 on the student website).

To protect declining industries. The human costs of sudden industrial closures can be very high. In such circumstances, temporary protection may be justified to allow the industry to decline more slowly, thereby avoiding excessive structural unemployment. Such policies will be at the expense of the consumer, however, who will be denied access to cheaper foreign imports. Nevertheless, such arguments have gained huge support from populist movements in the USA and elsewhere and protection for such industries form part of President Trump's 'America first' policies.

'Non-economic arguments'

A country may be prepared to forgo the direct economic advantages of free trade in order to achieve objectives that are often described as 'non-economic':

- It may wish to maintain a degree of self-sufficiency in case trade is cut off in times of war. This may apply particularly to the production of food and armaments.
- It may decide not to trade with certain countries with which it disagrees politically.
- It may wish to preserve traditional ways of life. Rural communities or communities built around old traditional industries may be destroyed by foreign competition.
- It may prefer to retain as diverse a society as possible, rather than one too narrowly based on certain industries.

Pursuing such objectives, however, will involve costs. Preserving a traditional way of life, for example, may mean that consumers are denied access to cheaper goods from abroad. Society must therefore weigh up the benefits against the costs of such policies.

Pause for thought

If economics is the study of choices of how to use scarce resources, can these other objectives be legitimately described as 'non-economic'?

BOX 14.3 DO WE EXPLOIT FOREIGN WORKERS BY BUYING CHEAP FOREIGN IMPORTS?

EXPLORING ECONOMICS

People sometimes question the morality of buying imports from countries where workers are paid 'pittance' wages. 'Is it right', they ask, 'for us to support a system where workers are so exploited?' As is often the case with emotive issues, there is some truth and some misunderstanding in a point of view like this.

First the truth. If a country like the UK trades with a regime that denies human rights and treats its workers very badly, then we may thereby be helping to sustain a corrupt system. We might also be seen to be lending it moral support. In this sense, therefore, trade may not help the cause of the workers in these countries. Arguments like these were used to support the imposition of trade sanctions against South Africa in the days of apartheid.

Now the misunderstanding. If we buy goods from countries that pay low wages, we are *not* as a result contributing to their low-wage problem. Quite the reverse. If countries like Indonesia export textiles to the West, this will help to *increase* the wages of Indonesian workers. If Indonesia has a comparative advantage in labour-intensive goods, these goods will earn a better price by being exported than by being sold entirely in the domestic Indonesian market. Provided *some* of the extra revenues go to the workers (as opposed to their bosses), they will gain from trade.

Under what circumstances would a gain in revenues by exporting firms not lead to an increase in wage rates?

Problems with protection

TC 1 p 7

Tariffs and other forms of protection impose a cost on society. This is illustrated in Figure 14.5, which shows the case of a good that is partly home produced and partly imported. Domestic demand and supply are given by D_{dom} and S_{dom}. It is assumed that firms in the country produce under perfect competition and that therefore the supply curve is the sum of the firms' marginal cost curves.

Let us assume that the country is too small to affect world prices: it is a price taker. The world price is given, at P_w. At P_w, Q_2 is demanded, Q_1 is supplied by domestic suppliers and hence $Q_2 - Q_1$ is imported.

Now a tariff is imposed. This increases the price to consumers by the amount of the tariff. Price rises to $P_w + t$. Domestic production increases to Q_3, consumption falls to Q_4, and hence imports fall to $Q_4 - Q_3$.

What are the costs of this tariff to the country? Consumers are having to pay a higher price, and hence consumer surplus falls from area *ABC* to *ADE* (see pages 201–2 if you are unsure about consumer surplus).

The cost to consumers in lost consumer surplus is thus *EDBC* (i.e. areas 1 + 2 + 3 + 4). *Part* of this cost, however, is redistributed as a *benefit* to other sections in society. *Firms* get a higher price, and thus gain extra producer surplus (area 1): where producer surplus is given by the area between the price and the *MC* curve. The *government* receives extra revenue from the tariff payments (area 3): i.e. $Q_4 - Q_3 \times$ tariff. These revenues can be used, for example, to reduce taxes.

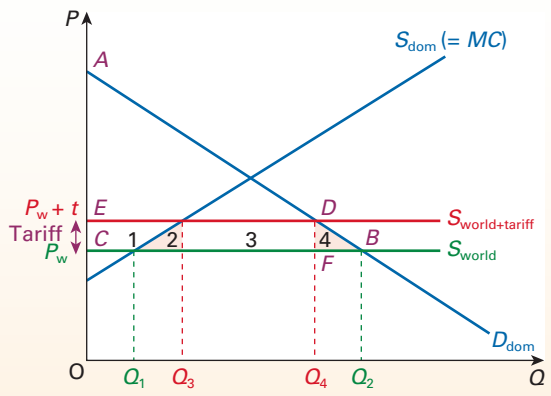
But *part* of this cost is not recouped elsewhere. It is a net cost to society (areas 2 and 4).

Area 2 represents the extra costs of producing $Q_3 - Q_1$ at home, rather than importing it. If $Q_3 - Q_1$ were still imported, the country would only be paying P_w. By producing it at home, however, the costs are given by the domestic supply curve (= *MC*). The difference between *MC* and P_w (area 2) is thus the efficiency loss on the production side.

Area 4 represents the loss of consumer surplus by the reduction in consumption from Q_2 to Q_4. Consumers have saved area FBQ_2Q_4 of expenditure, but have sacrificed area DBQ_2Q_4 of utility in so doing – a net loss of area 4.

The government should ideally weigh up such costs against any benefits that are gained from protection.

Figure 14.5 The cost of protection



Pause for thought

In this model, where the country is a price taker and faces a horizontal supply curve (the small country assumption), is any of the cost of the tariff borne by the overseas suppliers?

Apart from these direct costs to the consumer, there are several other problems with protection. Some are a direct effect of the protection; others follow from the reactions of other nations.

Protection as 'second-best'. Many of the arguments for protection amount merely to arguments for some type of government intervention in the economy. Protection, however, may not be the best way of dealing with the problem, since protection may have undesirable side effects. There may be a more direct form of intervention that has no side effects. In such a case, protection will be no more than a *second-best* solution.

Pause for thought

a. Protection to allow the exploitation of monopoly/monopsony power can be seen as a 'first-best' policy for the country concerned. Similarly, the use of tariffs to counteract externalities directly involved in the trade process (e.g. the environmental costs of an oil tanker disaster) could be seen to be a first-best policy. Explain why.

b. All the other arguments for tariffs or other forms of protection that we have considered can really be seen as arguments for intervention, with protection being no more than a second-best form of intervention. Go through each of the arguments and consider what would be a 'first-best' form of intervention.

For example, using tariffs to protect old inefficient industries from foreign competition may help prevent unemployment in those parts of the economy, but the consumer will suffer from higher prices. A better solution would be to subsidise retraining and investment in those areas of the country in *new efficient*

industries – industries with a comparative advantage. In this way, unemployment is avoided, but the consumer does not suffer.

TC 14 p 249

World multiplier effects. If a country, like the USA, imposes tariffs or other restrictions, imports will be reduced. But these imports are other countries' exports. A reduction in their exports will reduce the level of injections into the 'rest-of-the-world' economy, and thus lead to a multiplied fall in rest-of-the-world income. This in turn will lead to a reduction in demand for American exports. This, therefore, tends to undo the benefits of the tariffs.

Retaliation. If the USA imposes restrictions on, say, imports from the EU, then the EU may impose restrictions on imports from the USA. Any gain to US firms competing with EU imports is offset by a loss to US exporters. What is more, US consumers suffer, since the benefits from comparative advantage have been lost. (See, for example, Case Study 14.8 on the student website.)

The increased use of tariffs and other restrictions can lead to a trade war, with each country cutting back on imports from other countries. In the end, with such beggar-my-neighbour policies, everyone loses.

Protection may allow firms to remain inefficient. By removing or reducing foreign competition, protection may reduce firms' incentive to reduce costs. Thus if protection is being given to an infant industry, the government must ensure that the lack of competition does not prevent it 'growing up'. Protection should not be excessive and should be removed as soon as possible.

TC 6 p 24

Bureaucracy. If a government is to avoid giving excessive protection to firms, it should examine each case carefully. This can lead to large administrative costs. It could also lead to corrupt officials accepting bribes from importers to give them favourable treatment.

Recap

1. Countries use various methods to restrict trade, including tariffs, quotas, exchange controls, import licensing, export taxes, and legal and administrative barriers. Countries may also promote their own industries by subsidies.
2. Reasons for restricting trade that have some validity in a world context include the infant industry argument, the problems of relying on exporting goods whose market is growing slowly or even declining, dumping and other unfair trade practices, the danger of the establishment of a foreign-based monopoly, the need to spread the risks of fluctuating export prices, and the problems that free trade may adversely affect consumer tastes, that it may allow the importation of harmful goods and not take account of externalities.
3. Countries may also have other objectives in restricting trade, such as remaining self-sufficient in certain strategic products, not trading with certain countries of which it disapproves, protecting traditional ways of life or simply retaining a non-specialised economy.
4. Protection in the form of tariffs results in higher prices. The resulting loss in consumer surplus is not fully offset by the gain in profits to domestic firms and the tariff revenue for the government. Even if government intervention to protect certain parts of the economy is desirable, restricting trade is unlikely to be a first-best solution to the problem, since it involves side-effect costs. What is more, restricting trade may have adverse world multiplier effects; it may encourage retaliation; it may allow inefficient firms to remain inefficient; it may involve considerable bureaucracy and possibly even corruption.

14.4 THE WORLD TRADING SYSTEM AND THE WTO

Is trade becoming freer or less free?

After the Wall Street crash of 1929 (when prices on the US stock exchange plummeted), the world plunged into the Great Depression. Countries found their exports falling dramatically and many suffered severe balance of payments difficulties. The response of many countries was to restrict imports by the use of tariffs and quotas. Of course, this reduced other countries' exports, which encouraged them to resort to even greater protectionism. The net effect of the Depression and the rise in protectionism was a dramatic fall in world trade. The volume of world trade in manufactures fell by more than a third in the three years following the Wall Street crash. Clearly there was a net economic loss to the world from this decline in trade.

After the Second World War there was a general desire to reduce trade restrictions, so that all countries could gain the maximum benefits from trade.

There was no desire to return to the beggar-my-neighbour policies of the 1930s.

In 1947, 23 countries got together and signed the General Agreement on Tariffs and Trade (GATT). As of 2022 there were 164 members of its successor organisation, the World Trade Organization, which was formed in 1995. Between them, the members of the WTO account for over 98 per cent of world trade. The aims of GATT, and now the WTO, have been to liberalise trade.

WTO rules

The WTO requires its members to operate according to various principles. These include the following:

- *Non-discrimination*. Under the 'most favoured nations clause', any trade concession that a country makes to one member must be granted to *all* signatories. The only exception is with free-trade areas and customs unions (such as the EU). Here countries are permitted to abolish tariffs between themselves while still maintaining them with the rest of the world.
- *Reciprocity*. Any nation benefiting from a tariff reduction made by another country must reciprocate by making similar tariff reductions itself.
- *The general prohibition of quotas*.
- *Fair competition*. If unfair barriers are erected against a particular country, the WTO can sanction retaliatory action by that country. The country is not allowed, however, to take such action without permission.
- *Binding tariffs*. Countries cannot raise existing tariffs without negotiating with their trading partners.

Unlike the GATT, the WTO has the power to impose sanctions on countries breaking trade agreements. If there are disputes between member nations, these will be settled by the WTO, and if an offending country continues to impose trade restrictions, permission will be granted for other countries to retaliate.

For example, in March 2002, the Bush administration imposed tariffs on steel imports into the USA in order to protect the ailing US steel industry (see Case Study 14.6 on the student website). The EU and other countries referred the case to the WTO, which in December 2003 ruled that they were illegal. This ruling made it legitimate for the EU and other countries to impose retaliatory tariffs on US products. President Bush consequently announced that the steel tariffs would be abolished.

Then, following tariffs imposed by the Trump administration in early 2018 on Chinese steel and aluminium imports to the USA, China lodged a request with the WTO for consultations with the USA over the issue.

Pause for thought

Could US action to protect its steel industry from foreign competition be justified in terms of the interests of the USA as a whole (as opposed to the steel industry in particular)?

The greater power of the WTO has persuaded many countries to bring their disputes to it. From January 1995 to May 2022, 612 disputes had been brought to the WTO (compared with 300 to GATT over the whole of its 48 years).

Trade rounds

Periodically, member countries have met to negotiate reductions in tariffs and other trade restrictions. There have been eight 'rounds' of such negotiations since the signing of GATT in 1947. The last major round to be completed was the Uruguay Round, which began in Uruguay in 1986, continued at meetings around the world and culminated in a deal being signed in April 1994. By that time, the average tariff on manufactured products was 4 per cent and falling. In 1947 the figure was nearly 40 per cent. The Uruguay Round agreement also involved a programme of phasing in substantial reductions in tariffs and other restrictions up to the year 2002 (see Case Study 14.4 on the student website).

Despite the reduction in tariffs, many countries have still tried to restrict trade by various other means, such as quotas and administrative barriers. Also, barriers have been particularly high on certain non-manufactures. Agricultural protection in particular has come in for sustained criticism by developing countries. High fixed prices and subsidies given to farmers in the EU, the USA and other advanced countries mean that the industrialised world continues to export food to many developing countries which have a comparative advantage in food production! Farmers in developing countries often find it impossible to compete with subsidised food imports from the rich countries.

The most recent round of trade negotiations began in Doha, Qatar, in 2001 (see Box 14.4). The negotiations have focused on both trade liberalisation and measures to encourage development of poorer countries. In particular, the Doha Development Agenda, as it is called, is concerned with measures to make trade fairer so that its benefits are spread more evenly around the world. This would involve improved access for developing countries to markets in the rich world. The Agenda is also concerned with the environmental impacts of trade and development.

The negotiations were originally due to be completed in 2005, but, as Box 14.4 explains, deadlines continued to be missed. However, some progress was made at Ministerial Conferences in 2013 and 2015 as the box explains.

BOX 14.4 THE DOHA DEVELOPMENT AGENDA

CASE STUDIES & APPLICATIONS

A new direction for the WTO?

> Globalisation, based on the free play of comparative advantage, economies of scale and innovation, has produced a genuinely radical force, in the true sense of the word. It essentially amplifies and reinforces the strengths, but also the weaknesses, of market capitalism: its efficiency, its instability, and its inequality. If we want globalisation not only to be efficiency-boosting but also fair, we need more international rules and stronger multilateral institutions.[1]

In November 1999, the members of the World Trade Organization met in Seattle in the USA. What ensued became known as the 'Battle of Seattle' (see Case Study 14.7 on the student website). Anti-globalisation protesters fought with police; the world's developing economies fell out with the world's developed economies; and the very future of the WTO was called into question. The WTO was accused of being a free traders' charter, in which the objective of free trade was allowed to ride roughshod over anything that might stand in its way. Whatever the issue – the environment, the plight of developing countries, the dominance of trade by multinationals – free trade was king.

As Pascal Lamy, the former EU Trade Commissioner, made clear in the quote above, rules had to be strengthened, and the WTO had to ensure that the gains from trade were fairer and more sustainable.

The rebuilding process of the WTO began in Doha, Qatar, in 2001. The meeting between the then 142 members of the WTO concluded with the decision to launch a new round of WTO trade talks, to be called the 'Doha Development Agenda' (DDA). As with previous trade rounds, the talks were designed to increase the liberalisation of trade. However, this time such a goal was to be tempered by a policy of strengthening assistance to developing economies.

Other areas identified for discussion included: sustainable development and the environment; greater liberalisation of agriculture; rules to govern foreign direct investment; the co-ordination of countries' competition policies; and the use and abuse of patents on medicines.

The talks were originally scheduled for completion by January 2005, but this deadline was extended several times as new talks were arranged but failed to reach agreement. Rich countries were unwilling to make sufficient reductions in agricultural protection, given the pressure from their domestic farmers. The USA was unwilling to make substantial cuts in agricultural subsidies and the EU in agricultural tariffs.

There was also unwillingness by large developing countries, such as India and Brazil, to reduce protection to their industrial and service sectors. What is more, there were large divergences in opinion between developing countries on how much they should reduce their own agricultural protection.

Breakdown of the talks

The talks seemed finally to have broken down at a meeting in Geneva in July 2008. Despite the willingness of developing countries to reduce industrial tariffs by more than 50 per cent, and by the USA and the EU to make deep cuts in agricultural subsidies and tariffs, the talks foundered over the question of agricultural protection for developing countries. This was item 18 on a 'to-do' list of 20 items; items 1 to 17 had already been agreed. China and India wanted to protect poor farmers by retaining the ability to impose temporary tariffs on food imports in the event of a drop in food prices or a surge in imports. The USA objected. When neither side would budge, the talks collapsed.

Many commentators, however, argued that failure was no catastrophe. The gain from total liberalisation of trade would have boosted developing countries' GDP by no more than 1 per cent. And anyway, tariffs were generally falling and were already at an all-time low. But, with the global economic downturn of 2008/9, there were worries that protectionism would begin to rise again. This was a classic prisoners' dilemma (see pages 150–1 and 153). Policies that seemed to be in the interests of countries separately would be to the overall detriment of the world. The Nash equilibrium of such a 'game', therefore, is one where countries are generally worse off. As it turned out, the worries were largely unfounded – at least in the short term.

The Bali and Nairobi Packages

In December 2013, agreement was reached on a range of issues at the WTO's Bali Ministerial Conference and these were adopted in November 2014 by the General Council. The agreement means a streamlining of trade to make it 'easier, faster and cheaper', with particular focus on the promotion of development: boosting the trade of the least developed countries and allowing developing countries more options for providing food security, as long as this does not distort international trade.

This was the first significant agreement of the round and goes some way to achieving around 25 per cent of the goals set for the Doha Round.

Then in December 2015, at the Ministerial Conference in Nairobi, another historic agreement was made on various trade initiatives that should provide particular benefits to the WTO's poorest members. This 'Nairobi Package' contains six Ministerial Decisions on agriculture, cotton and issues related to least-developed countries, including a commitment to abolish export subsidies for farm exports. Such subsidies had been widely used by developed countries as a means of protecting their agricultural sector.

The end of the road?

Despite some progress, this could be the end of the road for the Doha Round. First, the Trump presidency was seen to reflect a new wave of protectionism, with free trade blamed for the decline of many traditional sectors, job losses and increased social deprivation. Second, many governments have pursued bilateral or plurilateral (multi-country) trade agreements. In contrast to multilateral WTO agreements, the result could be a global patchwork of trading rules, regulations and standards.

Does the process of globalisation mean that the role of the WTO is becoming less and less important?

Conduct a literature search around the topic of international trade and inequality. Summarise your findings in a PowerPoint presentation that could be presented to an audience of non-specialists in this area, and which would last for around 10–15 minutes.

[1] 'Global policy without democracy', speech by Pascal Lamy, given in 2001 when he was the EU Trade Commissioner. He later became head of the WTO in 2005.

Recap

1. Most countries of the world are members of the WTO and in theory are in favour of moves towards freer trade.
2. The WTO is more powerful than its predecessor, GATT. It has a disputes procedure and can enforce its rulings. In practice, however, countries have been very unwilling to abandon restrictions if they believe that they can gain from them, even though it might be at the expense of other countries.
3. WTO members periodically meet in rounds of talks. The latest, the Doha Round, aims to spread the benefits of trade across developing countries. It has yet to be concluded, but progress was made at ministerial conferences in 2013 and 2015. There is some doubt, however, as to whether any further progress will be made and whether the round, therefore, is effectively over. Many countries have moved to bilateral or plurilateral trading agreements.

14.5 TRADING BLOCS

Why do some countries get together and trade more freely between themselves?

The world economy has, over time, seen the formation of a series of trade blocs These have often been based upon regional groupings of countries, such as the European Union (EU) and the United States–Mexico–Canada Agreement (USMCA). Such trade blocs are examples of **preferential trading arrangements**. These arrangements involve trade restrictions with the rest of the world, and lower or zero restrictions between the members.

Although trade blocs clearly encourage trade between their members, many countries outside these blocs complain that they lose. For many developing economies, in need of access to the most prosperous nations in the world, this represents a significant check on their ability to grow and develop.

Types of preferential trading arrangement

There are three possible forms of such trading arrangements.

Free-trade areas

A **free-trade area** is where member countries remove tariffs and quotas between themselves, but retain whatever restrictions *each member chooses* with non-member countries. Some provision will have to be made to prevent imports from outside coming into the area via the country with the lowest external tariff. An example is the EU–UK Trade and Cooperation Agreement (see Section 14.7).

Customs unions

A **customs union** is like a free-trade area, but in addition members must adopt *common* external tariffs and quotas with non-member countries.

Common markets

A **common market** is where member countries operate as a *single* market. Like a customs union, there are no tariffs and quotas between member countries and there are common external tariffs and quotas. But a full common market includes additional features:

- *A common system of taxation*. In the case of a *perfect* common market, this will involve identical rates of tax in all member countries.
- *A common system of laws and regulations governing production, employment and trade*. For example, in a perfect common market there would be a *single* set of laws governing issues such as product specification (e.g. permissible artificial additives to foods, or levels of exhaust emissions from cars), the employment and dismissal of labour, mergers and takeovers, and monopolies and restrictive practices.
- *Free movement of labour, capital and materials, and of goods and services*. In a perfect common market, this will involve a total absence of border controls between member states, the freedom of workers to work in any member country, and the freedom of firms to expand into any member state.
- *The absence of special treatment by member governments of their own domestic industries*. Governments are large purchasers of goods and services. In a perfect common market, they should buy from whichever companies within the market offer the most competitive deal and not show favouritism towards domestic suppliers: they should operate a *common procurement policy*.

Definitions

Preferential trading arrangement A trading arrangement whereby trade between the signatories is freer than trade with the rest of the world.

Free-trade area A group of countries with no trade barriers between themselves.

Customs union A free-trade area with common external tariffs and quotas.

Common market A customs union where the member countries act as a single market with free movement of labour and capital, common taxes and common trade laws.

The definition of a common market is sometimes extended to include the following two features of *economic and monetary union*.

- *A fixed exchange rate between the member countries' currencies.* In the extreme case, this would involve a single currency for the whole market.
- *Common macroeconomic policies.* To some extent this must follow from a fixed exchange rate, but in the extreme case it will involve a single macroeconomic management of the whole market, and hence the abolition of separate fiscal or monetary intervention by individual member states.

We will examine European economic and monetary union in Section 15.6.

The direct effects of a customs union: trade creation and trade diversion

By joining a customs union (or free-trade area), a country will find that its trade patterns change. Two such changes can be distinguished: trade creation and trade diversion.

Trade creation

Trade creation is where consumption shifts from a high-cost producer to a low-cost producer. The removal of trade barriers allows greater specialisation according to comparative advantage. Instead of consumers having to pay high prices for domestically produced goods in which the country has a comparative disadvantage, the goods can now be obtained more cheaply from other members of the customs union. In return, the country can export to them goods in which it has a comparative advantage.

KI 36 p 422

Trade diversion

Trade diversion is where consumption shifts from a lower-cost producer outside the customs union to a higher-cost producer within the union.

Assume that the most efficient producer in the world of a particular good is Russia – outside the EU. Assume that before membership, Poland paid a similar tariff on this good from any country, and thus imported the product from Russia rather than from the EU.

TC 1 p 7

After joining the EU, however, the removal of the tariff made the EU product cheaper, since the tariff remained on the Russian product. Consumption thus switched to a higher-cost producer. There was thus a net loss in world efficiency.

As far as Poland was concerned, consumers still gained, since they were paying a lower price than before. There was a loss, however, to domestic producers (from the reduction in protection, and hence reduced prices and profits) and to the government (from reduced tariff revenue). These losses may have been smaller or larger than the gain to consumers: in other words, there may still have been a net gain to Poland, but there could have been a net loss, depending on the circumstances. See Web Appendix 14.3 on the student website for a graphical analysis of these arguments.

Pause for thought

Is joining a customs union more likely to lead to trade creation or trade diversion in each of the following cases? (a) The union has a very high external tariff. (b) Cost differences are very great between the country and members of the union.

Longer-term effects of a customs union

Over the longer term, there may be other gains and losses from being a member of a customs union.

Longer-term advantages

- Increased market size may allow a country's firms to exploit (*internal*) *economies of scale*. This argument is more important for small countries, which have therefore more to gain from an enlargement of their markets.
- *External economies of scale.* Increased trade may lead to improvements in the infrastructure of the members of the customs union (better roads, railways, financial services, etc.). This in turn could bring bigger long-term benefits from trade between members, and from external trade too, by making the transport and handling of imports and exports cheaper.

KI 28 p 196

- The bargaining power of the whole customs union with the rest of the world may allow member countries to gain *better terms of trade*. This, of course, will necessarily involve a degree of political co-operation between the members.

KI 19 p 122

- *Increased competition* between member countries may stimulate efficiency, encourage investment and reduce monopoly power. Of course, a similar advantage could be gained by the simple removal of tariffs with any competing country.
- Integration may encourage a *more rapid spread of technology.*

Longer-term disadvantages

- *Resources may flow from the country* to more efficient members of the customs union, or to the

Definitions

Trade creation Where a customs union leads to greater specialisation according to comparative advantage and thus a shift in production from higher-cost to lower-cost sources.

Trade diversion Where a customs union diverts consumption from goods produced at a lower cost outside the union to goods produced at a higher cost (but tariff free) within the union.

geographical centre of the union (so as to minimise transport costs). This can be a major problem for a common market (where there is free movement of labour and capital). The country could become a depressed 'region' of the community.

- If integration encourages greater co-operation between firms in member countries, it may also encourage *greater oligopolistic collusion*, thus keeping prices higher to the consumer. It may also encourage mergers and takeovers that would increase monopoly power. The extent to which this takes place will depend on whether the customs union has a joint competition policy and how effective it is.
- *Diseconomies of scale*. If the union leads to the development of very large companies, they may become bureaucratic and inefficient.
- The *costs of administering* the customs union may be high. This problem is likely to be worse the more intervention there is in the affairs of individual members.

Preferential trading in practice

Preferential trading has the greatest potential to benefit countries whose domestic market is too small, taken on its own, to enable them to benefit from economies of scale, and where they face substantial barriers to their exports. Most developing countries fall into this category and as a result many have attempted to form preferential trading arrangements.

Examples in Latin America and the Caribbean include the Latin American Integration Association (ALADI), the Andean Community, the Central American Integration System (SICA) and the Caribbean Community (CARICOM). A Southern Common Market (MerCoSur) was formed in 1991, consisting of Argentina, Brazil, Paraguay and Uruguay. Venezuela joined in 2012 but its membership was suspended in 2016. MerCoSur has a common external tariff and most of its internal trade is free of tariffs.

The Association of South-East Asian Nations (ASEAN) was formed in 1967 when six nations (Brunei, Indonesia, Malaysia, the Philippines, Singapore and Thailand) agreed to work towards an ASEAN Free Trade Area (AFTA). Between 1984 and 1999 they were joined by four new members (Vietnam, Lao, Myanmar and Cambodia). ASEAN has a population of over 650 million people and is dedicated to increased economic co-operation within the region.

By 2010, virtually all tariffs between the six original members had been eliminated and both tariff and non-tariff barriers are falling quickly for both original and new members. The ASEAN Economic Community (AEC) was established in 2015, ahead of schedule.

In Africa, the Economic Community of West African States (ECOWAS) has been attempting to create a common market between its 15 members which has a combined population of around 350 million. The West African franc is used in eight of the countries and another six plan to introduce a common currency, the eco. After several delays, a roadmap was agreed in 2021 for a launch of the eco in 2027. The ultimate goal is to combine the two currency areas and adopt a single currency for all member states.

United States–Mexico–Canada Agreement (USMCA)

The USMCA is a free-trade agreement between the USA, Mexico and Canada. It is the successor to the North American Free Trade Agreement (NAFTA) and, alongside the EU, is one of the two most powerful trading blocs in the world. NAFTA came into force in 1994 when the three countries agreed to abolish tariffs between themselves in the hope that increased trade and co-operation would follow. Tariffs between the USA and Canada were phased out by 1999, and between all three countries by January 2008. Many non-tariff restrictions remain, although, under the original agreement, new ones were not to be permitted.

The election of Donald Trump as President was a pivotal moment for NAFTA. During his campaign, Trump had described NAFTA as the 'worst trade deal in US history'. Shortly after coming to office, he committed to renegotiating NAFTA. These negotiations began in August 2017. They were often difficult and, in June 2018, the Trump administration imposed tariffs on steel and aluminium imports of 25 per cent and 10 per cent respectively from several countries, including Canada and Mexico. However, a deal was eventually reached in September 2018, with the new USMCA being implemented from April 2020 following ratification by the three countries.

There were three main changes. First, there are stricter rules-of-origin requirements for motor vehicles. To qualify for zero tariffs, 75 per cent of their components must be made within USMCA (not the previous 62.5 per cent), with at least 40 to 45 per cent of a vehicle's components being made by workers earning at least $16 per hour (by 2023). Second, Canada agreed to give US dairy farmers access to 3.6 per cent of Canada's dairy market. Third, various standards were strengthened, such as in digital trade and financial services.

While USMCA has a market size similar to that of the EU, it is at most only a free-trade area and not a common market. Unlike the EU, it does not seek to harmonise laws and regulations, except in very specific areas such as environmental management and labour standards. Member countries are permitted total legal independence, subject to the one proviso that they must treat firms of other member countries equally with their own firms – the principle of 'fair

competition'. Nevertheless, NAFTA and then USMCA have encouraged a growth in trade between its members, most of which is trade creation rather than trade diversion.

Case Study 14.9 on the student website looks at the costs and benefits of NAFTA/USMCA membership for the three countries involved.

Asia–Pacific Economic Cooperation (APEC)

The most significant move towards establishing a more widespread regional economic organisation in east Asia appeared with the creation of the Asia-Pacific Economic Co-operation (APEC) in 1989. APEC links 21 economies of the Pacific rim, including Asian, Australasian and North and South American countries (19 countries, plus Hong Kong and Taiwan). These countries account for over half of global GDP and almost half of the world's trade. At the 1994 meeting of APEC leaders, it was resolved to create a free-trade area across the Pacific by 2010 for the developed industrial countries, and by 2020 for the rest.

This preferential trading area is by no means as advanced as NAFTA/USMCA and is unlikely to move beyond a free-trade area. Within the region there exists a wide disparity across a range of economic and social indicators. Such disparities create a wide range of national interests and goals. Countries are unlikely to share common economic problems or concerns. In addition, political differences and conflicts within the region are widespread, reducing the likelihood that any organisational agreement beyond a simple economic one would succeed.

The Trans-Pacific Partnership

The Trans-Pacific Trade Partnership (TPP) agreement was signed in 2016 by 12 Pacific-rim countries, including the USA but not China. But on coming into office in 2017, Donald Trump withdrew the USA from the agreement. In 2018 the remaining TPP countries signed a largely unchanged version known as the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP). By 2021, the agreement had come into force in 8 of the 11 countries.

The CPTPP is more than a simple free-trade agreement. In terms of trade it involves the removal of many non-tariff barriers as well as most tariff barriers. It also has elements of a single market. For example, it contains many robust and enforceable environmental protection, human rights and labour standards measures. It also allows for the free transfer of capital by investors in most circumstances.

However, it also established an 'investor–state dispute settlement' (ISDS) mechanism. This allows companies from any of the TPP countries to sue governments of any other countries in the agreement for treaty violations, such as giving favourable treatment to domestic companies, the seizing of companies' assets, or controls over the movement of capital. Critics of ISDS claim that it gives too much power to companies and may prevent governments from protecting their national environment or domestic workers and companies.

In 2021 the UK began the process of formally applying to join the CPTPP. Other countries, including South Korea, Indonesia, the Philippines, Thailand, Taiwan and Colombia, have also expressed an interest in joining. Whether the USA will eventually join remains to be seen.

In the next section we consider the longest established and most comprehensive preferential trading arrangement: the European Union.

Recap

1. Countries may make a partial movement towards free trade by the adoption of a preferential trading system. This involves free trade between the members, but restrictions on trade with the rest of the world. Such a system can be either a simple free-trade area, or a customs union (where there are common restrictions with the rest of the world) or a common market (where in addition there is free movement of capital and labour, and common taxes and trade laws).
2. A preferential trading area can lead to trade creation, where production shifts to low-cost producers within the area, or to trade diversion, where trade shifts away from lower-cost producers outside the area to higher-cost producers within the area.
3. Preferential trading may bring longer-term advantages of increased economies of scale (both internal and external), improved terms of trade from increased bargaining power with the rest of the world, increased efficiency from greater competition between member countries and a more rapid spread of technology. On the other hand, it can lead to increased regional problems for members, greater oligopolistic collusion and various diseconomies of scale. There may also be large costs of administering the system.
4. There have been several attempts around the world to form preferential trading systems. The two most powerful are the European Union and the United States–Mexico–Canada Agreement (USMCA).

14.6 THE EUROPEAN UNION

What have been the effects of the creation of a 'single market' in the EU?

The European Economic Community (EEC) was formed by the signing of the Treaty of Rome in 1957 and came into operation on 1 January 1958.

The original six member countries of the EEC (Belgium, France, Italy, Luxembourg, Netherlands and West Germany) had already made a move towards integration with the formation of the European Coal and Steel Community in 1952. This had removed all restrictions on trade in coal, steel and iron ore between the six countries. The aim had been to gain economies of scale and allow more effective competition with the USA and other foreign producers.

The EEC extended this principle and aimed eventually to be a full common market with completely free trade between members in all products, and with completely free movement of labour, enterprise and capital.

All internal tariffs between the six members had been abolished and common external tariffs established by 1968. But this still only made the EEC a *customs union*, since a number of restrictions on internal trade remained (legal, administrative, fiscal, etc.). Nevertheless, the aim was eventually to create a full common market.

In 1973 the UK, Denmark and Ireland became members. Greece joined in 1981, Spain and Portugal in 1986, and Sweden, Austria and Finland in 1995, Then in May 2004 a further ten countries joined: Cyprus, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia and Slovenia. Bulgaria and Romania joined in 2007. The last new member is Croatia, which joined in 2013. When the UK left the EU in 2020, there were then 27 members.

From customs union to common market

The EU is clearly a customs union. It has common external tariffs and no internal tariffs. But is it also a common market? For many years there have been *certain* common economic policies.

Common Agricultural Policy (CAP). The Union has traditionally set common high prices for farm products. This has involved charging variable import duties to bring foreign food imports up to EU prices and intervention to buy up surpluses of food produced within the EU at these above-equilibrium prices. Although the main method of support has shifted to providing subsidies (or 'income support') unrelated to current output, this still represents a common economic policy of agricultural support.

Regional policy. EU regional policy provides grants to firms and local authorities in relatively deprived regions of the Union.

Competition policy. EU policy here has applied primarily to companies operating in more than one member state. For example, Article 101 of the Treaty of Lisbon prohibits agreements between firms operating in more than one EU country (e.g. over pricing or sharing out markets) which adversely affect competition (see Case Studies 6.18 and 6.20 on the student website).

Harmonisation of taxation. VAT is the standard form of indirect tax throughout the EU. There are, however, substantial differences in VAT rates between member states, as there are with other tax rates (see Box 14.5).

Social policy. In 1989 the European Commission presented a *Social Charter* to the EU heads of state. This spelt out a series of worker and social rights that should apply in all member states (see Case Study 14.11 on the student website). These rights were grouped under 12 headings covering areas such as the guarantee of decent levels of income for both the employed and the non-employed, freedom of movement of labour between EU countries, freedom to belong to a trade union and equal treatment of women and men in the labour market. The Social Charter was only a recommendation and each element had to be approved separately by the European Council of Ministers.

KI 24 p 185

The Social Chapter of the Maastricht Treaty (1991) attempted to move the Community forward in implementing the details of the Social Charter in areas such as maximum hours, minimum working conditions, health and safety protection, information and consultation of workers, and equal opportunities.

Pause for thought

Would the adoption of improved working conditions necessarily lead to higher labour costs per unit of output?

Despite these various common policies, in other respects the Community of the 1970s and 1980s was far from a true common market: there were all sorts of non-tariff barriers such as high taxes on wine by non-wine producing countries, special regulations designed to favour domestic producers, governments giving contracts to domestic producers (e.g. for defence

BOX 14.5 FEATURES OF THE SINGLE MARKET

CASE STUDIES & APPLICATIONS

Since 1 January 1993 trade within the EU has operated very much like trade within a country. In theory, there should be no more difficulty for a firm in Marseilles to sell its goods in Berlin than in Paris. At the same time, the single market allows free movement of labour and involves the use of common technical standards.

The features of the single market are summed up in two European Commission publications.[1] They are:

- Elimination of border controls on goods within the EU: no more long waits.
- Free movement of people across borders.
- Common security arrangements.
- No import taxes on goods bought in other member states for personal use.
- The right for everyone to live in another member state.
- Recognition of vocational qualifications in other member states: engineers, accountants, medical practitioners, teachers and other professionals able to practise throughout Europe.
- Technical standards brought into line, and product tests and certification agreed across the whole EU.
- Common commercial laws – making it attractive to form Europe-wide companies and to start joint ventures.
- Public contracts to supply equipment and services to state organisations now open to tenders across the EU.

So, what does the single market mean for individuals and for businesses?

Individuals

Before 1993, if you were travelling in Europe, you had a 'duty-free allowance'. This meant that you could only take goods up to the value of €600 across borders within the EU without having to pay VAT in the country into which you were importing them. Now you can take as many goods as you like from one EU country to another, provided they are for your own consumption. But to prevent fraud, member states may ask for evidence that the goods have been purchased for the traveller's own consumption if they exceed specified amounts.

Individuals have the right to live and work in any other member state. Qualifications obtained in one member state must be recognised by other member states.

Firms

Before 1993 all goods traded in the EU were subject to VAT at every internal border. This involved some 60 million customs clearance documents resulting in a cost of some €70 per consignment.[2]

This has all now disappeared. Goods can cross from one member state to another without any border controls: in fact, the concepts of 'importing' and 'exporting' within the EU no longer officially exist. All goods sent from one EU country to another will be charged VAT only in the country of destination. They are exempt from VAT in the country where they are produced.

One of the important requirements for fair competition in the single market is the convergence of tax rates. Although income tax rates, corporate tax rates and excise duties still differ between member states, there has been some narrowing in the range of VAT rates. Higher rates of VAT on luxury goods were abolished and countries are allowed to have no more than two lower rates of at least 5 per cent on 'socially necessary' goods, such as food and water supply.

There is now a lower limit of 15 per cent on the standard rate of VAT. Standard rates in January 2022 nonetheless varied from 17 per cent in Luxembourg to 27 per cent in Hungary. However, during the early 2010s several countries, including the UK, Ireland, Greece, Italy, Hungary and Portugal, increased their standard rate of VAT as a means of reducing their budget deficits (see Section 13.1). One effect of this has been that the vast majority of EU countries now have a standard rate of VAT between 20 and 25 per cent.

In what ways would competition be 'unfair' if VAT rates differed widely between member states?

Using the Public Finances Databank from the Office for Budget Responsibility, create a chart showing for the UK the percentage of public-sector current receipts collected from VAT (net of refunds) over time. Briefly summarise the findings of your chart.

[1] *A Single Market for Goods* (Commission of the European Communities, 1993); *10 Key Points about the Single European Market* (Commission of the European Communities, 1992).

[2] See *A Single Market for Goods* (Commission of the European Communities, 1993).

equipment), and so on. The Single European Act of 1986, however, sought to remove these barriers and to form a genuine common market by the end of 1992 (see Box 14.5).

The benefits and costs of the single market

It is difficult to quantify the benefits and costs of the single market, given that many occur over a long period, and that it is difficult to know to what extent the changes that are taking place are the direct result of the single market.

In 2012, the European Commission published *20 Years of the European Single Market*. This stated that, 'EU27 GDP in 2008 was 2.13 per cent or €233 billion higher than it would have been if the Single Market had not been launched in 1992. In 2008 alone, this amounted to an average of €500 extra in income per person in the EU27. The gains come from the Single Market programme, liberalisation in network industries such as energy and telecommunication, and the enlargement of the EU to 27 member countries.'

Even though the precise magnitude of the benefits is difficult to estimate, it is possible to identify the

types of benefit that have resulted, many of which have been substantial.

KI 36 p 422

Trade creation. Costs and prices have fallen as a result of a greater exploitation of comparative advantage. Member countries can now specialise further in those goods and services that they can produce at a comparatively low opportunity cost.

Reduction in the direct costs of barriers. This category includes administrative costs, border delays and technical regulations. Their abolition or harmonisation has led to substantial cost savings.

Economies of scale. With industries based on a Europe-wide scale, many firms can now be large enough, and their plants large enough, to gain the full potential economies of scale (see Box 5.6 on page 108). Yet the whole European market is large enough for there still to be adequate competition. Such gains have varied from industry to industry, depending on the minimum efficient scale of a plant or firm. Economies of scale have also been gained from mergers and other forms of industrial restructuring

KI 20 p 127

Greater competition. Increased competition between firms has led to lower costs, lower prices and a wider range of products available to consumers. This has been particularly so in newly liberalised service sectors such as transport, financial services, telecommunications and broadcasting. In the long run, greater competition can stimulate greater innovation, the greater flow of technical information and the reorganisation ('rationalisation') of production to increase efficiency.

Despite these gains, the single market has not received a universal welcome within the EU. Its critics argue that, in a Europe of oligopolies, unequal ownership of resources, rapidly changing technologies and industrial practices, and factor immobility, the removal of internal barriers to trade has merely exaggerated the problems of inequality and economic power. More specifically, the following criticisms are made.

Radical economic change is costly. Substantial economic change is necessary to achieve the full economies of scale and efficiency gains from a single European market. These changes necessarily involve redundancies – from bankruptcies, takeovers, rationalisation and the introduction of new technology. The severity of this 'structural' and 'technological' unemployment (see Section 12.1) depends on (a) the pace of economic change and (b) the mobility of labour – both occupational and geographical. Clearly, the more integrated markets become across the EU, the less the costs of future change.

Adverse regional effects. Firms are likely to locate as near as possible to the 'centre of gravity' of their markets and sources of supply. If, before barriers are removed, a firm's prime market was the UK, it might well have located in the Midlands or the north of England. If, however, with barriers now removed, its market has become Europe as a whole, it may choose to locate in the south of England or in France, Germany or the Benelux countries instead. The creation of a single European market thus tends to attract capital and jobs away from the edges of the Union and towards its geographical centre.

In an ideal market situation, areas like the south of Italy or Portugal should attract resources from other parts of the Union. Being relatively depressed areas, wage rates and land prices are lower. The resulting lower industrial costs should encourage firms to move into the areas. In practice, however, as capital and labour (and especially young and skilled workers) leave the extremities of the Union, so these regions are likely to become more depressed. If, as a result, their infrastructure is neglected, they then become even less attractive to new investment.

TC 14 p 249

The development of monopoly/oligopoly power. The free movement of capital can encourage the development of giant 'Euro-firms' with substantial economic power. Indeed, recent years have seen some very large European mergers (see Case Study 6.21 on the student website). This can lead to higher, not lower prices, and less choice for the consumer. It all depends on just how effective competition is, and how effective EU competition policy is in preventing monopolistic and collusive practices.

KI 19 p 122

Trade diversion. Just as increased trade creation has been a potential advantage of completing the internal market, so trade diversion has been a possibility too. This is more likely if *external* barriers remain high (or are even increased) and internal barriers are *completely* abolished.

Perhaps the biggest objection raised against the single European market is a political one: the loss of national sovereignty. Governments find it much more difficult to intervene at a microeconomic level in their own economies. This was one of the key arguments in the debate over Britain's future within Europe in the run-up to the EU referendum in 2016 (see Section 14.6).

Completing the internal market

Despite the reduction in barriers in the 1990s, the internal market is still not 'complete'. In other words, various barriers to trade between member states still remain.

To monitor progress what is now known as the Single Market Scoreboard was established in 1997. This shows progress towards the total abandonment of any forms of internal trade restrictions (Case Study 14.13 on the student website). It shows

the percentage of EU Single Market Directives still to be transposed into national law: the 'transposition deficit'. To counteract new barriers, the EU periodically issues new Directives. If this process is more rapid than that of the transposition of existing Directives into national law, the transposition deficit increases.

In 1997, the average transposition deficit of member countries was 6.3 per cent. By 1999, this had fallen to 3.5 per cent and to 0.5 per cent in 2014. The normal range is now between 0.5 and 0.7 per cent, although it occasionally rises above this if there are a lot of new directives waiting to be transposed.

The Scoreboard also identifies the number of infringements of the internal market that have taken place. The hope is that the 'naming and shaming' of countries will encourage them to make more rapid progress towards totally free trade within the EU.

Despite the general success in reducing the transposition deficits, national governments have continued to introduce *new* technical standards, several of which have had the effect of erecting new barriers to trade. Also, infringements of single market rules by governments have not always been dealt with. The net result is that, although trade is much freer today than in the early 1990s, especially given the transparency of pricing with the euro, various barriers still exist, especially to the free movement of goods.

The effect of the new member states

Given the very different nature of the economies of many of the newer entrants to the EU, and their lower levels of GDP per head, the potential for gain from membership has been substantial. The gains come through trade creation, increased competition, technological transfer and inward investment, both from other EU countries and from outside the EU.

A study in 2004[2] concluded that Poland's GDP would rise by 3.4 per cent and Hungary's by almost 7 per cent. Real wages would rise, with those of unskilled workers rising faster than those of skilled workers, in accordance with these countries' comparative advantage. There would also be benefits for the then 15 existing EU countries' members from increased trade and investment, but these would be relatively minor in comparison to the gains to the new members.

A European Commission Report[3] produced in April 2009, five years after the enlargement, found that the expansion had been a win–win situation for both old and new members. There had been significant improvements in the standard of living in new member states and they had benefited from modernisation of their economies and more stabilised institutions and laws. In addition, enterprises in old member states had enjoyed opportunities for new investment and exports, and there had been an overall increase in trade and competition between the member states.

Pause for thought

Why may the new members of the EU have the most to gain from the single market, but also the most to lose?

In future years, now that the euro is used by at least 20 of the member states, with the possibility of others joining at some time, trade within the 27 EU countries is likely to continue to grow as a proportion of GDP. We examine the benefits and costs of the single currency and the whole process of economic and monetary union in the EU in Section 15.6.

[2] M. Maliszewska *Benefits of the Single Market expansion for current and new member states* (Centrum Analiz Społeczno-Ekonomicznych, 2004).

[3] 'Five Years of an Enlarged EU – Economic Achievements and Challenges', *European Economy 1 2009* (Commission of the European Communities).

Recap

1. The European Union is a customs union, in that it has common external tariffs and no internal ones. But virtually from the outset it has also had elements of a common market, particularly in the areas of agricultural policy, regional policy, competition policy, and to some extent in the areas of tax harmonisation and social policy.
2. Nevertheless, there were substantial non-tariff barriers to trade within the EU. The Single European Act of 1986 sought to sweep away these restrictions and to establish a genuine free market within the EU: to establish a full common market. Benefits from completing the internal market have included trade creation, cost savings from no longer having to administer barriers, economies of scale for firms now able to operate on a Europe-wide scale, and greater competition leading to reduced costs and prices, greater flows of technical information and more innovation.
3. Critics of the single market point to the costs of radical changes in industrial structure, the attraction of capital away from the periphery of the EU to its geographical centre, possible problems of market power with the development of giant 'Euro-firms', and the possibilities of trade diversion.
4. The actual costs and benefits of EU membership to the various countries vary with their particular economic circumstances. These costs and benefits in the future will depend on just how completely the barriers to trade are removed, on the extent of monetary union and on the effects of enlarging the Union.

14.7 THE UK AND BREXIT

What impact will new trading arrangements have for the UK?

On 23 June 2016, the UK held a referendum on whether to remain a member of the EU. 72.1 per cent of the electorate voted and by a majority of 51.9 per cent to 48.1 per cent, Britain voted to leave the EU.

Alternative trading arrangements

The focus then turned to the alternative trading arrangements that could be agreed by the UK and EU. Table 14.1 helps to understand the options.

One possibility was the *Norwegian model*. This would have seen the UK join the European Economic Area (EEA), giving it access to the single market, but removing regulation in some key areas, such as fisheries and home affairs. This was ruled out in favour of a *bilateral agreement*. Three main types were available:

- *Canadian model* where the UK would form a comprehensive trade agreement with the EU to lower customs tariffs and other barriers to trade.
- *Swiss model* where the UK would negotiate a series of bilateral agreements with the EU, including selective or general access to the single market.
- *Turkish model* where the UK would form a customs union with the EU. In Turkey's case the agreement relates principally to manufactured goods.

The government's preference was for a comprehensive trade agreement. On the 31 January 2020, after several delays, the UK left the EU. There then followed a transition period up to 31 December 2020, during which the UK remained in the single market and the customs union, while negotiations on a trade agreement took place.

On 24 December 2020, EU and UK negotiators reached an agreement on the text of the EU–UK Trade and Cooperation Agreement (TCA). This took effect on 1 January 2021, when the UK became a 'third country': a non-EU country, where its citizens no longer enjoyed the rights to free movement across EU member states.

The EU–UK TCA is a variant of the Canadian model, providing for zero tariffs and quotas on all goods. However, for tariff-free access, businesses must show that their products fulfil rules-of-origin requirements. Under these rules, when a good is imported into the UK from outside the EU and then has value added to it by processing, packaging, remixing, preserving, etc., it will only count as a UK good if sufficient value or weight is added. The proportions vary by product, but typically goods must have at least 50 per cent UK content (or 80 per cent of the weight in the case of foodstuffs).

The TCA does not include free trade in services. UK service providers thus face new barriers, which impose costs. For example, financial services firms have lost *passporting rights*, which allowed them to sell their services into the EU without the need for additional regulatory clearance. Also, *mutual recognition* for service providers no longer automatically applies, thus affecting the recognition of some UK qualifications.

The Protocol on Ireland and Northern Ireland means that Northern Ireland is subject to some EU rules related to the single market for goods and the customs union. Checks and controls take place on goods entering Northern Ireland from the rest of the UK, and EU customs duties then apply to those goods deemed at risk of moving on to the EU.

Long-term growth, trade and Brexit

The effects of Brexit will take many years to become clear. Many of these effects will depend on the impact that the decision has for the UK's long-term rate of economic growth.

In Chapter 10 we saw that long-term growth requires continuous increases in potential output. Furthermore, when analysing the effect on living standards, it is the growth in real GDP *per capita* that is important (see page 286). Hence, a crucial part of the longer-term analysis of Brexit is its impact on the supply-side of the economy, including, therefore its impact on labour productivity, capital accumulation, human capital development and technological progress.

The important point here is that for most countries it is important to consider their long-term growth and well-being within the context of a global economy with markedly greater levels of openness and trade flows than in the past.

TC 15 p 287

Since the UK joined the EU in 1973 the openness of the UK economy has increased rapidly. Figure 14.6 shows the ratio of the total flow of import and exports of goods and services to GDP since the 1960s. From 1960 to 2020, it nearly trebled from just over 20 per cent to just over 60 per cent (before falling somewhat during the pandemic).

Table 14.2 Alternative trading relationships with the EU

Trading arrangement		Tariffs	Customs union & external trade	Non-tariff barriers /other policy and regulatory issues
EU membership		Full tariff-free trade	Common external tariffs No customs costs Access to EU Free Trade Agreements (FTAs)	Alignment of regulations, standards and specifications Non-discriminatory access for markets for services
EEA (Norway)		Some tariffs on agriculture and fisheries	Custom costs apply No access to EU Free Trade Agreements (FTAs)	Limited coverage of agricultural and fisheries Compliance with most EU rules and standards, including free movement of people and social policy
Bilateral agreements	TCA	Zero tariffs and zero quotas on all goods	Custom costs apply No access to EU free-trade agreements (FTAs)	No financial services passport or mutual recognition for service suppliers (e.g. of personal qualifications) Compliance with EU standards for firms importing into EU
	Switzerland	Some tariffs on agriculture	Custom costs apply No access to EU Free Trade Agreements (FTAs)	Minimises non-tariff barriers in areas covered by agreements Limited coverage of services No financial services passport Complies with EU rules in sector covered by agreements, including free movements of people and social policy
	Canada	Some tariffs on agriculture Tariffs for transitional period on manufactured goods	Custom costs apply No access to EU Free Trade Agreements (FTAs)	No financial services passport Compliance with EU standards for firms importing into EU
	Turkey	Tariff exemptions apply only to manufactured goods and processed agricultural goods	No custom costs for manufactured goods Align external trade policy with EU	No financial services passport No special access for services Adopts EU product standards Compliance with environmental standards linked to goods and to rules on competition and state aid
WTO membership		EU external tariffs apply	Custom costs apply No access to EU Free Trade Agreements (FTAs)	No financial services passport Compliance with EU standards for firms importing into EU

Adapted from *EU Referendum: HM Treasury analysis key facts,* HM Treasury (18 April 2016) (available at https://www.gov.uk/government/news/eu-referendum-treasury-analysis-key-facts)

Adverse effects of Brexit

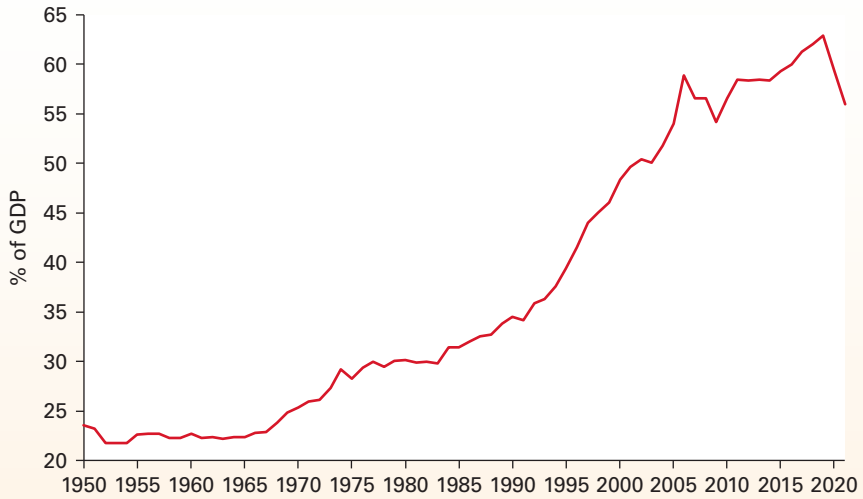
In its 2016 analysis of the possible long-term implications of Brexit, the UK Treasury argued that the country's openness to trade and investment had been a key factor behind the growth in the economy's potential output.[4] Hence, maintaining this openness, it argued, would be important for long-term growth and so for raising living standards. Positive supply-side effects from openness might include:

- Increasing market opportunities which enable firms to exploit internal economies of scale (see pages 102–3).
- Increasing competition encourages firms to improve their productivity to maintain their market share and so encourages the adoption of new technologies and processes.
- Knowledge and technology transfer. The diffusion of ideas and technological know-how can be passed between firms in international supply chains, through internationally mobile workers or new international entrants.

The Treasury's analysis estimated the average impact of Brexit on households by modelling the adverse impacts on the supply-side of the economy from lower levels of openness. To do so it estimated the impact 15 years after leaving the EU of the various possible types of trading arrangement shown in Table 14.2. It found that GDP would be between

[4] *EU Referendum: HM Treasury analysis key facts*, HM Treasury (18 April 2016) (available at https://www.gov.uk/government/news/eu-referendum-treasury-analysis-key-facts).

Figure 14.6 UK trade in goods and services as a percentage of GDP



Notes: Exports *(X)* plus imports *(M)* as a percentage of GDP; all series at constant prices

Source: Based on data from ONS *Time Series Data* (series IKBK, IKBL, ABMI)

3.4 per cent and 9.5 per cent lower, depending on the deal agreed.

The Treasury estimated that with a Norwegian-type deal, households would be between £2400 and £2900 worse off each year at 2015 prices after 15 years. With a negotiated bilateral trade agreement they would be between £3200 and £5400 worse off, while no deal and reversion to WTO trading terms would create a household loss per annum of between £3700 and £6600 at 2015 prices.

The OECD suggested that Brexit would be like a tax, pushing up the costs and weakening the economy. Its analysis indicated that by 2030, GDP would be at least 5 per cent lower than otherwise, causing households to be £3200 worse off (at 2016 prices). It continued that:

> In the longer term, structural impacts would take hold through the channels of capital, immigration and lower technical progress. In particular, labour productivity would be held back by a drop in foreign direct investment and a smaller pool of skills. The extent of forgone GDP would increase over time . . . The effects would be even larger in a more pessimistic scenario and remain negative even in the optimistic scenario.[5]

The OECD analysis points to the *structural change* the UK economy is likely to experience. The growth in openness experienced by the UK economy has occurred within the context of EU membership. Membership has influenced the patterns of trade and investment. It has provided the framework in which businesses have operated – for example, the development of supply chains across EU member countries. Structural changes accompanying Brexit, the OECD argued, would have negative supply-side effects.

Pause for thought

Can we incorporate trade effects into endogenous growth models?

Opportunities from Brexit

Despite the pessimistic forecasts from the vast majority of economists about a British exit, there was a group of eight economists in favour of Brexit.[6] They claimed that leaving the EU would lead to a stronger economy, with higher GDP, a faster growth in real wages, lower unemployment and a smaller gap between imports and exports. The main argument to support the claims was that the UK would be more able to pursue trade creation freed from various EU rules and regulations. In other words, there would be positive supply-side effects.

In reply, many economists argue that the EU has much greater bargaining power to achieve more favourable trade deals than the UK could by acting alone.

[5] *The economic consequences of Brexit: a taxing decision*, OECD (25 April 2016) (available at http://www.oecd.org/economy/the-economic-consequences-of-brexit-a-taxing-decision.htm).

[6] www.economistsforbrexit.co.uk (site now discontinued).

While disagreement about the impact of the UK's exit from the EU was to be expected, there was agreement that these effects would work primarily through their impact on the supply-side.

TC 15 p 287

However, perhaps less clear are the likely *distributional* effects of the UK's exit. Trade can impact on different sectors differently. For example, people working in agriculture, often on low incomes, will be affected by the replacement of the Common Agricultural Policy. Thus, the distributional effects will depend on the new funding arrangements for the sector and the support schemes put in place.

Consequently, a more complete analysis of trading relationships and hence of Brexit must take account of distributional effects.

KI 31 p 206

Pause for thought

What sectors might a distributional analysis of the impact of trade consider?

Recap

1. Following a referendum in June 2016 the UK voted to leave the European Union. On 31 January 2020, the UK left the EU, at which point began an 11-month transition period and negotiations on a new trade agreement.
2. On 24 December 2020, negotiators reached an agreement – a new EU–UK Trade and Cooperation Agreement (TCA). This provides for zero tariffs and zero quotas on all goods. However, to benefit, businesses must fulfil rules-of-origin requirements. The agreement does not include free trade in services.
3. Economists have largely argued that Brexit will result in negative supply-side effects by impeding cross-border trade and reducing the UK's openness to trade and investment. This adversely affects productivity growth and living standards.
4. Some economists, however, have argued that outside of the EU the UK would be free of EU rules and regulations and able to create trade. The supply-side benefits would raise long-term growth and living standards.
5. Trade and the openness of economies have distributional effects. A more complete economic assessment of Brexit therefore requires that we consider such effects.

14.8 TRADE AND DEVELOPING COUNTRIES

The importance of international trade to developing countries

The role of international trade is one of the most contentious for developing countries. Should they adopt an open trading policy with few if any barriers to imports? Should they go further and actively promote trade by subsidising their export sector? Or should they restrict trade and pursue a policy of greater self-sufficiency? These are issues we shall be examining in this section.

Whether it is desirable that developing countries should adopt policies of more trade or less, trade is still vital. Certain raw materials, capital equipment and intermediate products that are necessary for development can be obtained only from abroad. Others *could* be produced domestically but only at much higher cost.

The relationship between trade and development

What makes the issue of trade so contentious is the absence of a simple relationship between trade and development. Instead, the relationship is complex and determined by a series of interactions between variables affecting both trade and development. Furthermore, while some countries have managed to use trade as an engine for economic growth and wider human development, others, despite trade liberalisation, have seen relatively little improvement in either their export performance or in human development.

In constructing a trade and development index, the UN[7] identified three broad groups of influences or dimensions which interact and affect a country's trade and development performance. Within these broad groups are various indicators which themselves interact. We consider briefly these three dimensions and some of the indicators within each dimension.

Structural and institutional dimension

Human capital. This relates to the skills and expertise of the workforce which affect a country's performance and its productivity (see Section 10.5). Education and

[7] *Trade and development index*, United Nations (https://unctad.org/system/files/official-document/ditctab20051ch1_en.pdf).

health are key influences here. As well as affecting the economic growth of a country, higher educational attainment and better health conditions positively impact on social and human development.

Physical infrastructure. Infrastructure affects a country's productive capacity and so its potential output. Poor transport infrastructure, for example, is thought to be a major impediment on a country's export performance.

Financial environment. Credit is important to producers and consumers alike in helping to finance both short-term and longer-term commitments. For instance, it enables firms to finance day-to-day operational purchases but also longer-term investments in fixed assets such as buildings and machinery.

Institutional quality. This relates to issues of governance not just of firms themselves, but also of institutions, largely governmental.

Environment sustainability. The argument here is that excessive activity, particularly at the early stages of development, can result in environmental degradation. This can adversely affect human development and, in turn, economic development.

Trade policies and processes dimension

Openness to trade. In the absence of market failures and externalities, trade liberalisation is argued to be a driver of development. However, there can be significant human costs in the transition process.

Effective access to foreign markets. The success of a country's export performance is crucially dependent on its access to markets. Barriers to access include tariffs and non-tariff barriers, such as regulatory standards in the markets of recipient countries. A wider definition of 'effective' access recognises other factors too. These might include the size of foreign markets, transport links, the characteristics of the goods being exported – for example how differentiated they are – as well as the cost of the exported goods.

Levels of development dimension

The third series of factors affecting both trade and human development relate to existing levels of development. We can think of the relevant development issues here as encompassing three components: economic development, social development and gender development.

Trade strategies

Despite the complexity of the relationship between trade and development, countries' policies towards trade typically go through various stages as they develop.

Primary outward-looking stage. Traditionally, developing countries have exported primaries – minerals such as copper, cash crops such as coffee, and non-foodstuffs such as cotton – in exchange for manufactured consumer goods. Having little in the way of an industrial base, if they want to consume manufactured goods, they have to import them.

Secondary inward-looking stage. In seeking rapid economic development, most developing countries drew lessons from the experience of the advanced countries. The main conclusion was that industrialisation was the key to economic success.

But industrialisation required foreign exchange to purchase capital equipment. This led to a policy of **import-substituting industrialisation**, which involved cutting back on non-essential imports and thereby releasing foreign exchange. Tariffs and other restrictions were imposed on those imports for which a domestic substitute existed or which were regarded as unimportant.

Secondary outward-looking stage. Once an industry had satisfied domestic demand, it had to seek markets abroad if expansion was to continue. What is more, as we shall see, import substitution brought a number of serious problems for developing countries. The answer seemed to be to look outward again, this time to the export of manufactured goods. Many of the most economically successful developing countries have owed their high growth rates to a rapid expansion of manufactured exports. Recent examples include China, India and Indonesia, while countries like Hong Kong, Singapore, South Korea, Taiwan have now progressed to the status of 'advanced economies'.

We will now examine the three stages in more detail.

Approach 1: Exporting primaries – exploiting comparative advantage

The justification for exporting primaries

Despite moves towards import substitution and secondary export promotion, many developing countries still rely heavily on primary exports. Three major arguments have traditionally been used for pursuing a policy of exporting primaries. In each case the arguments have also been used to justify a policy of free or virtually free trade.

Definition

Import-substituting industrialisation (ISI) A strategy of restricting imports of manufactured goods and using the foreign exchange saved to build up domestic industries.

Exporting primaries exploits comparative advantage. Traditional trade theory implies that countries should specialise in producing those items in which they have a comparative advantage: i.e. those goods that can be produced at relatively low opportunity costs. For most developing countries this means that a large proportion of their exports should be primaries.

KI 36 p 422

Exporting primaries provides a 'vent for surplus'. Trade offers a vent for surplus: i.e. a means of putting to use resources that would otherwise not be used. These surpluses occur where the domestic market is simply not big enough to consume all the available output of a particular good. There is far too little demand within Zambia to consume its potential output of copper. The same applies to Namibian uranium and Peruvian tin.

Exporting primaries provides an 'engine for economic growth'. According to this argument, developing countries benefit from the growth of the economies of the developed world. As industrial expansion takes place in the rich North, this creates additional demand for primaries from the poor South. In more recent years, the rapid growth in China, India and other industrialising developing countries has seen a rapid growth in demand for commodities, many produced in the least developed countries. This helps to drive up commodity prices (e.g. during the 2000s), which benefits primary exporters (see Figure 14.7 below). However, as we shall see, commodity prices can be very volatile (see also Box 2.3); indeed, they generally fell across the 2010s.

TC 14 p 249

Traditional trade theory in the context of development

There are several reasons for questioning whether the above arguments justify a policy of relying on primary exports as the means to development.

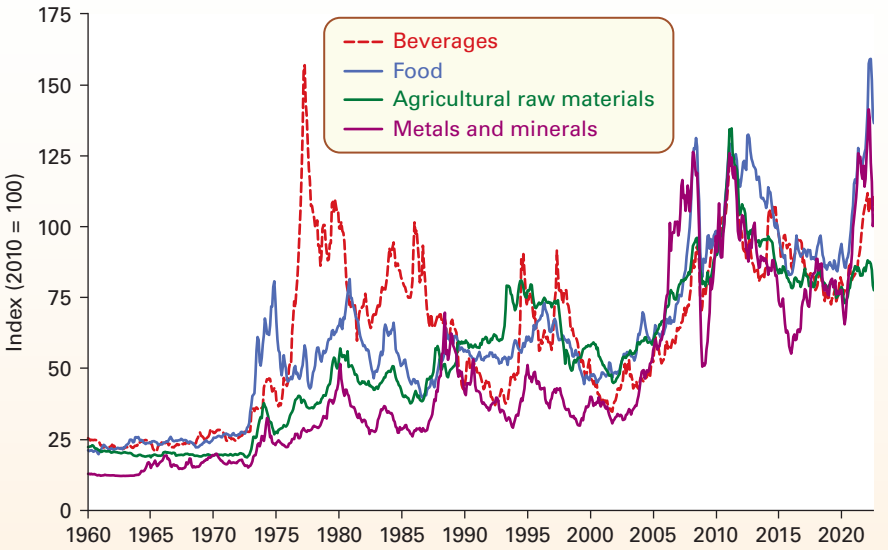
Comparative costs change over time. Over time, with the acquisition of new skills and an increase in the capital stock, a developing country that once had a comparative advantage in primaries may find that it now has a comparative advantage in certain *manufactured* products, especially those which are more labour intensive and use raw materials of which the country has a plentiful supply. The market, however, cannot necessarily be relied upon to bring about a smooth transition to producing such products.

KI 36 p 422

Concentrating on primary production may hinder growth. A country may gain in the short run by exporting primaries and using the money earned to buy imports. But the production of certain primaries may have little potential for expansion and thus growth may be slower.

The benefits from trade may not accrue to the nationals of the country. If a mine or plantation is owned by a foreign company, it will be the foreign shareholders who get the profits from the sale of exports. In addition, these companies may bring in their own capital and skilled labour from abroad. The benefits gained by the local people will probably be confined to the additional wages they earn. With these companies being in a position of monopsony power, these wages are often very low.

Figure 14.7 World primary commodity prices (2010 = 100)



Source: Based on data from *World Bank Commodity Price Data (The Pink Sheet)* (Commodity Markets, World Bank)

In recent years organisations like the Fair Trade movement have focused attention on the low incomes received by many primary producers and their lack of market power. Only a tiny fraction of the price you pay for coffee, tea or bananas goes to local growers.

KI 24 p 185

Trade may lead to greater inequality. Trade shifts income distribution in favour of those factors of production employed intensively in the export sector. If exports are labour-intensive, greater equality will tend to result. But if they are land- or raw material-intensive then trade will redistribute income in favour of large landowners or mine owners.

KI 28 p 196

Exporting primary exports may involve external costs. Mining can lead to huge external costs, such as the despoiling of the countryside and damage to the health of miners. Mines and plantations can lead to the destruction of traditional communities and their values.

These arguments cast doubt on whether a policy of relying on free trade in primary exports is the best way of achieving economic development. Various trends in the international economy have also worked against primary exporters.

Problems for primary exporters: long term

Long-term trends in international trade have caused problems for primary exporting countries in a number of ways.

TC 10 p 49

Low income elasticity of demand for certain primary products. As world incomes grow, so a smaller proportion of these incomes is spent on primaries. Since food is a necessity, consumers, especially in rich countries, already consume virtually all they require. A rise in incomes, therefore, tends to be spent more on luxury goods and services, and only slightly more on basic foodstuffs. The exceptions are certain 'luxury' imported foodstuffs such as exotic fruits. In the case of raw materials, as people's incomes grow, they tend to buy more and more expensive products. The extra value of these products, however, arises not from the extra raw materials they might contain, but from their greater sophistication.

This argument however has not always applied in recent years, however. The rapid growth of countries such as China and India, where people spend a relatively large proportion of any increase in their income on food, has led to periods of rapidly rising world food prices (see Box 2.3). This has been aggravated by poor harvests, switching land to grow crops for biofuels, and other supply disruptions: e.g. from the Russian invasion of Ukraine in 2022. This has raised concerns among developing countries about 'food security' and the ability of people to afford foodstuffs.

There has also been a rapid growth in demand by such countries for raw materials as inputs into the construction industry and the expanding industrial sector.

Agricultural protection in advanced countries. Faced with the problem of a slowly growing demand for food produced by their own farmers, advanced countries increasingly imposed restrictions on imported food. Reducing these restrictions has been one of the main aims of the Doha Development Agenda (the latest round of WTO trade negotiations: see Box 14.4).

Technological developments. Synthetic substitutes have in many cases replaced primaries in the making of consumer durables (such as furniture and household appliances), clothing and industrial equipment. Also, the process of miniaturisation, as microchips have replaced machines, has meant that less and less raw material has been required to produce any given amount of output.

TC 10 p 49

Rapid growth in imports. There tends to be a high income elasticity of demand for imported manufactures. This is the result partly of the better-off in developing countries increasingly being able to afford luxury goods, and partly of the development of new tastes as people are exposed to the products of the developed world – products such as Coca-Cola, Levi® jeans, mobile phones and iPods. In fact, the whole process has been dubbed 'Coca-Colanisation'. Because of a lack of domestic substitutes, the price elasticity of demand for manufactured imports is low. This gives market power to the overseas suppliers of these imports, which tends to raise their price relative to exports.

The terms of trade. Between 1980 and 2000 the prices of many primary products declined. For instance, the nominal price index in 2000 for beverages (coffee and tea) was only one-quarter of its level in the late 1970s – and even less than that in real terms. This reflected the slow growth in demand for primaries and led to a decline in the terms of trade for primary exporters. This is because they were having to export more and more in order to buy any given quantity of imports, such as manufactured goods.

As Figure 14.7 shows, a quite different picture emerged in the 2000s. As the demand for food and raw materials grew rapidly, reflecting the rapid growth of China and various other emerging economies, so primary commodity prices rose sharply.

This was to come to an abrupt halt with the world recession of 2008–9, when primary product prices fell sharply. The resumption of global growth saw primary commodity prices climb once more. Thereafter, global economic growth rates fluctuated and commodity prices fell. During the COVID-19 pandemic, the aggregate commodity price index fell to levels seen during the financial crisis. However, as economies began to open up and demand recovered, pandemic-related

supply disruptions and then the Russian invasion of Ukraine – both countries being significant exporters of energy and agricultural commodities – caused large spikes and volatility in commodity prices.

Approach 2: Import-substituting industrialisation (ISI)

Dissatisfaction with relying on primary exporting led most developing countries to embark on a process of industrialisation. The newly industrialised countries (NICs), such as China, Malaysia, Brazil and India, are already well advanced along the industrialisation road. Other developing countries have not yet progressed very far, especially the poorest African countries.

The most obvious way for countries to industrialise was to cut back on the import of manufactures and substitute them with home-produced manufactures. This could not be done overnight: it had to be done in stages, beginning with assembly, then making some of the components, and finally making all, or nearly all, of the inputs into production. Most developing countries have at least started on the first stage. Several of the more advanced developing countries have component manufacturing industries. Only a few of the larger NICs, such as China, India and Brazil, have built extensive capital goods industries.

The method most favoured by policy makers was **tariff escalation**. Here tariff rates (or other restrictions) increased as one moves from the raw materials to the intermediate product to the finished product stage. Thus finished goods had higher tariffs than intermediate products. This encouraged assembly plants, which were protected by high tariffs from imported finished products, and were able to obtain components at a lower tariff rate.

One of the problems with ISI was that countries were desperately short of resources to invest in industry. As a result, a policy of ISI usually involved encouraging investment by multinational companies. But even without specific 'perks' (e.g. tax concessions, cheap sites, the cutting of red tape), multinationals would still probably be attracted by the protection afforded by the tariffs or quotas.

Adverse effects of import substitution

Some countries, such as South Korea and Taiwan, pursued an inward-looking ISI policy for only a few years. For them it was merely a stage in development, rapidly to be followed by a secondary outward-looking policy. Infant industries were initially given protection, but when they had achieved sufficient economies of scale, the barriers to imports were gradually removed.

The countries that continued to pursue protectionist ISI policies generally had a poorer growth record. They also tended to suffer from other problems, such as a deepening of inequality. The development of the modern industrial sector was often to the detriment of the traditional sectors and also of the export sector.

The criticisms of ISI are numerous, including the following.

It ran directly counter to the principle of comparative advantage. Rather than confining ISI to genuine infant industries and then gradually removing the protection, ISI was applied indiscriminately to a whole range of industries. Countries found themselves producing goods in which they had a comparative *disadvantage*.

KI 36 p 422

It cushioned inefficient practices and encouraged the establishment of monopolies. Without the competition from imports, many of the industries were highly inefficient and wasteful of resources. What is more, in all but the largest or most developed of the developing countries the domestic market for manufactures is small. If a newly established industry is to be large enough to gain the full potential economies of scale, it must be large relative to the market. This means that it will have considerable monopoly power.

KI 19 p 122

It involved artificially low real interest rates. In order to encourage capital investment in the import-substituting industries, governments often intervened to keep interest rates low. This encouraged the use of capital-intensive technology with a consequent lack of jobs. It also starved other sectors (such as agriculture) of much needed finance, and discouraged saving.

It led to urban wages above the market-clearing level. Wage rates in the industrial sector, although still low compared with advanced countries, are often considerably higher than in the traditional sectors.

- They are pushed up by firms seeking to retain labour in which they have invested training.
- Governments, seeking to appease the politically powerful urban industrial working class, have often passed minimum wage laws.
- Trade unions, although less widespread in developing than in advanced countries, are mainly confined to the new industries.

Higher industrial wages again encourage firms to use capital-intensive techniques.

It involved overvalued exchange rates. Restricting imports tends to lead to an appreciation of the exchange rate. This makes non-restricted imports

Definition

Tariff escalation The system whereby tariff rates increase the closer a product is to the finished stage of production.

cheaper, which then discourages the production of domestic goods, such as food and component parts, which compete with those imports. Also, a higher exchange rate discourages exports. Exports tend to be priced in dollars. If the exchange rate appreciates, domestic currency will buy more dollars; or put another way, a dollar will exchange for less domestic currency. Thus exporters will earn less domestic currency as the exchange rate appreciates.

Pause for thought

Why is an overvalued exchange rate likely to encourage the use of capital-intensive technology?

It did not necessarily save on foreign exchange. Many of the new industries were highly dependent on the importation of raw materials, capital equipment and component parts. These imported inputs, unlike foreign finished goods, were often supplied by a single firm, which could thus charge monopoly prices. What is more, a large proportion of the extra incomes generated by these industries tended to be spent on imports by the new urban elites.

Protection was not applied evenly. Many different tariff rates were used in one country: in fact, a policy of tariff escalation demands this. In addition, governments often used a whole range of other protectionist instruments – such as the licensing of importers, physical and value quotas and foreign exchange rationing. These were often applied in a haphazard way. The result was that protection was highly uneven.

KI 24 p 185

Income distribution was made less equal. Additional incomes generated by the modern sector tended to be spent on modern-sector goods and imported goods. Thus there was a multiplier effect *within* the modern sector, but virtually none between the modern sector and the traditional sectors. Also, as we saw above, an overvalued exchange rate leads to a bias against agriculture and thus further deepens the divide between rich and poor. Finally, the relatively high wages of the modern sector encourage workers to migrate to the towns, where many, failing to get a job, live in dire poverty.

KI 28 p 196

Social, cultural and environmental problems. A policy of ISI often involves imposing an alien set of values. Urban life can be harsh, competitive and materialistic. Moreover, a drive for industrialisation may involve major costs to the environment, as a result of waste products from new industries, often with low environmental standards.

Finally, import substitution is necessarily limited by the size of the domestic market. Once that is saturated, ISI can come to an abrupt halt. At that stage, further expansion can only come from exporting; but if these industries have been overprotected, they are less able to compete successfully in world markets.

Approach 3: Exporting manufactures – a possible way forward?

The countries with the highest rates of economic growth have been those that have successfully made the transition to being exporters of manufactures. Table 14.3 gives some examples.

The transition from inward-looking to outward-looking industrialisation

How is a country to move from import substituting to being outward looking? One approach is to take it industry by industry. When an industry has saturated the home market and there is no further scope for import substitution, it should then be encouraged to seek markets overseas. The trouble with this approach is that if the country is still protecting other industries, there will probably still be an overvalued exchange rate. Thus specific subsidies, tax concessions or other 'perks' would have to be given to this industry to enable it to compete. The country would still be highly interventionist, with all the distortions and misallocation of resources this tends to bring.

The alternative is to wean the whole economy off protection. Three major things will need doing:

- a devaluation of the currency (see Section 15.4) in order to restore the potential profitability of the export sector;
- a dismantling of the various protective measures that had biased production towards the home market;
- a removal, or at least a relaxing, of price controls.

But these are things that cannot be done 'at a stroke'. Firms may have to be introduced gradually to the greater forces of competition that an outward-looking trade policy brings. Otherwise there may be massive bankruptcies and a corresponding massive rise in unemployment.

The benefits from a secondary outward-looking policy

The advocates of outward-looking industrialisation make a number of points in its favour.

It conforms more closely to comparative advantage. Countries pursuing an open trade regime will be able to export only those goods in which they have a comparative advantage. The resources used in earning a unit of foreign exchange from exports will be less than those used in saving a unit of foreign exchange by replacing imports with home-produced goods. In other words, resources will be used more efficiently.

KI 36 p 422

Table 14.3 Macroeconomic indicators of selected secondary outward-looking countries

	Average 1985–2020				2020	
	Annual growth in real GDP (%)	Annual growth in real GDP per capita (%)	Exports (% of GDP)	Exports of manufactures (% of merchandise exports)	GDP per capita, PPP, $	GNY per capita, PPP, $
Brazil	2.3	1.0	12	46	15 388	14 980
China	9.2	8.3	21	85	16 653	16 610
Hong Kong	3.8	2.9	154	73	62 106	65 240
India	5.8	4.1	15	70	6 998	6 930
Malaysia	5.4	3.2	86	66	29 623	28 850
Singapore	5.6	3.5	184	74	102 573	90 320
South Korea	5.7	5.0	36	90	42 998	43 370
Low-income economies	3.2	0.4	27	n/a	2 067	2 014
Middle-income economies	4.6	3.2	24	67	12 059	11 882
Low and middle-income economies	4.6	3.0	24	67	11 048	10 884
High-income economies	2.3	1.6	25	72	52 239	52 559
World	3.0	1.6	24	70	17 548	17 461

Source: Based on data from *Databank*, World Bank (2022)

BOX 14.6 THE EVOLVING COMPARATIVE ADVANTAGE OF CHINA

Riding the dragon

Comparative advantage enables specialisation and trade and this can be one of the key factors that helps a country to grow and develop. During the 1990s and 2000s, China experienced an average rate of growth of just over 10 per cent per year. Its emergence as an economic power is due to many things, but its ability to exploit its *comparative advantage* is certainly one such factor.

A country's comparative advantage often derives from its abundant resources and China used its abundance of cheap labour. With labour costs estimated to be between 60 and 90 per cent lower than in the USA, it was this that attracted many manufacturing companies to China, making a range of products using low- and moderately skilled jobs.

According to the 2022 World Investment Report, China (excluding Hong Kong) was the world's second largest *recipient* of foreign direct investment (FDI) with investment of $181 billion, the equivalent of 11.4 per cent of global FDI.

However, it is not just the quantity of labour that explained China's dominance in manufacturing. While many of its workers are low-skilled, many are educated. Furthermore, Paul Krugman notes another key factor:

> China's dominant role in the export of many labor-intensive manufactured goods surely reflects its combination of relatively abundant labor and relatively high manufacturing competence.[1]

It is unsurprising that multinational companies would take advantage of lower costs of production and locate factories in China. But, while China has benefited from this, many developed nations have seen a decline in their manufactured exports. Countries like the UK and USA gradually adjusted and moved to exploit their comparative advantage in the services sector. They saw a comparative disadvantage emerge in manufactured items. This changing comparative advantage as a country develops is well documented. Could it be that China will soon begin to see its own comparative advantage change?

Rising labour costs. China's low labour costs are crucial, but these have been rising, as workers demand higher wages, shorter hours and larger benefits. Data suggest that labour costs have been growing at some 20 per cent per year.

One effect of this has been for some labour-intensive businesses to migrate towards inland China, where labour costs are lower.

However, in other cases the move has been more significant. With a comparative advantage in low-cost labour disappearing, some labour-intensive businesses have left China, moving to other nations which can boast cheap labour, such as Bangladesh, Cambodia, Indonesia and Vietnam. This is especially the case for companies specialising in the production of clothes and shoes.

But the story will not stop there. These industries require labour, and as long as this remains the case, when a country begins to grow this will lead to higher wage demands, which in turn will raise costs. Production will shift once more.

Moving up the value chain

So, what does this mean for China? When the USA and Europe lost their comparative advantage in the production

[1] Paul Krugman, 'Increasing Returns in a Comparative Advantage World', in Robert M. Stern, *Comparative Advantage, Growth, and the Gains from Trade and Globalization* (World Scientific, 2011) Chapter 7, p. 45.

Economies of scale. If the home market is too small to allow a firm to gain all the potential economies of scale, these can be gained by expanding into the export market.

KI 20 p 127

Increased competition. By having to compete with foreign companies, exporters will be under a greater competitive pressure than industries sheltering behind protective barriers. This will encourage (a) resource saving in the short run through reductions in inefficiency, and (b) innovation and investment, as firms attempt to adopt the latest technology, often obtained from developed countries.

Increased investment. To the extent that outward-looking policies lead to a greater potential for economic growth, they may attract more foreign capital. To the extent that they involve an increase in interest rates, they will tend to encourage saving. To the extent that they lead to increased incomes, additional saving will be generated, especially given that the *marginal* propensity to save may be quite high. The extra savings can be used to finance extra investment.

It can lead to more employment and a more equal distribution of income. The manufactured goods in which a country has a comparative advantage will be those produced by labour-intensive techniques (since wage rates are comparatively low in developing countries). Export expansion will thus increase the demand for labour relative to capital and thus create more employment. The increased demand for labour will tend to lead to a rise in wages relative to profits.

KI 24 p 185

It removes many of the costs associated with ISI. Under a policy of ISI, managers may spend a lot of their time lobbying politicians and officials, seeking licences (and sometimes paying bribes to obtain them), adhering to norms and regulations or trying to find ways around them. If an outward-looking policy involves removing all this, managers can turn their attention to producing goods more efficiently.

TC 6 p 24

Drawbacks of an export-orientated industrialisation strategy

The export of manufactures is seen by many developed countries as very threatening to their own industries.

CASE STUDIES & APPLICATIONS

of manufactured products, they had to look elsewhere. They developed a comparative advantage in the production of products requiring highly skilled labour and increasingly specialised in the services sector. However, with rising costs in emerging economies, more manufacturing, especially of high-value products, is being returned to these developed nations.

China will need to follow the pattern of the Western economies; its companies and workers will need to move up the value chain to find products that they can specialise in, which are not easily transferable to lower-wage countries.[2] This is no easy task, as producing more high-value products will involve entering into direct competition with countries that have had time to develop their comparative advantage.

> China's reliance on cheap labor has powered the country's economy to unprecedented heights. But China's manufacturing sector is running into problems these days: squeezed from one end by places with even lower labor costs, such as Laos and Vietnam, and yet struggling to move to higher ground making more advanced products because of competition from developed nations such as Germany and the United States.[3]

In recent years, China has invested heavily in high-tech sectors, especially those such as solar energy, where demand is expected to grow rapidly. Since 2020, it has seen growth rates in excess of 100 per cent per year in products such as new-energy vehicles, industrial robots, excavation machinery, transportation and microcomputer equipment. With a huge domestic market, a large R&D sector to support such developments and a large planned sector to provide predictable demand, China faces far less risk than smaller market economies heavily reliant on the export sector.

The changing nature of comparative advantage can cause problems for workers, businesses and wider society, but benefits also emerge through greater competition, choice and innovation. The key is how countries adapt. This is important because most countries can expect to experience further waves of industrial change.

1. *Why are countries likely to see their comparative advantage change as they develop?*
2. *In what ways does a booming Chinese economy benefit the rest of the world?*

Using the UNCTADStat Data Center download data on inward flows of FDI into China as a percentage of global FDI flows. Plot the data in a line chart and then briefly summarise the findings of your chart.

[2] Mohan Kompella, 'China, comparative advantage and moving up the value chain', *The Story of Business blog* (25 November 2012).

[3] Jia Lynn Yang, 'China's manufacturing sector must reinvent itself, if it's to survive', *The Washington Post* (23 November 2012).

Their response has often been to erect trade barriers. These barriers have tended to be highest in the very industries (such as textiles, footwear and processed food) where developing countries have the greatest comparative advantage. Even if the barriers are *currently* low, developing countries may feel that it is too risky to expand their exports of these products for fear of a future rise in barriers. Recognising this problem, the WTO is very keen to ensure fair access for developing countries to the markets of the rich world. This has been a core focus of the Doha Round of trade negotiations (see Box 14.4).

The successes of developing countries such as China, India and Malaysia in exporting manufactures do not imply that other developing countries will have similar success. As additional developing countries attempt to export their manufactures, they will be facing more and more competition from each other.

Another problem is that, if a more open trade policy involves removing or reducing exchange and capital controls, the country may become more vulnerable to speculative attack. This was one of the major factors contributing to the east Asian crisis of the late 1990s. Gripped by currency and stock market speculation, and by banking and company insolvency, many countries of the region found that economic growth had turned into a major recession. The 'miracle' seemed to be over. Nevertheless, the countries with the least distortions fared the best during the crisis. Thus Singapore and Taiwan, which are open and relatively flexible, experienced only a slowdown, rather than a recession.

Exporting manufactures can be a very risky strategy for developing countries, such as many in Africa. Perhaps the best hope for the future lies in a growth in manufacturing trade *between* developing countries. That way they can gain the benefits of specialisation and economies of scale that trade brings, while at the same time producing for a growing market. The feasibility of this approach depends on whether developing countries can agree to free-trade areas or even customs unions (see Section 14.5).

There does, however, seem to be a strong movement in this direction (see pages 433–4). The share of trade between developing countries (so-called South–South trade) in total world exports has doubled over the past 20 years, to over 25 per cent. Manufactured goods now account for nearly 60 per cent of South–South trade.

Pause for thought

Why may small developing countries have more to gain than large ones from a policy of becoming exporters of manufactured products?

Recap

1. Trade is of vital importance for the vast majority of developing countries, and yet most developing countries suffer from chronic balance of trade deficits.
2. Developing countries have traditionally been primary exporters. This has allowed them to exploit their comparative advantage in labour-intensive goods and has provided a market for certain goods that would otherwise have no market at home.
3. There are reasons for questioning the wisdom of relying on traditional primary exports, however. With a low world income elasticity of demand for many primary products, with the development of synthetic substitutes for minerals and with the protection of agriculture in developed countries, the demand for primary exports from the developing world has grown only slowly. At the same time, the demand for manufactured imports into developing countries has grown rapidly. Until recently, the result was a decline in the terms of trade. The rapid growth in demand for primary products from China and India has, to some extent, reversed this trend and the terms of trade have improved for many primary exporters.
4. Import-substituting industrialisation (ISI) was seen to be the answer to the problems of primary exporting. ISI was normally achieved in stages, beginning with the finished goods stage and then working back towards the capital goods stage. ISI, it was hoped, would allow countries to benefit from the various long-term advantages associated with manufacturing.
5. For many countries, however, ISI brought as many problems, if not more, than it solved. It often led to the establishment of inefficient industries, protected from foreign competition and facing little or no competition at home either. It led to considerable market distortions, with tariffs and other forms of protection haphazardly applied; to overvalued exchange rates, with a resulting bias against exports and the agricultural sector generally; to a deepening of inequalities and to large-scale social problems as the cities expanded, as poverty and unemployment grew and as traditional values were undermined; to increased dependency on imported inputs; and to growing environmental problems.
6. The most rapidly growing of the developing countries are those that have pursued a policy of export-orientated industrialisation. This has allowed them to achieve the benefits of economies of scale and foreign competition, and to specialise in goods in which they have a comparative advantage (i.e. labour-intensive goods) and yet which have a relatively high income elasticity of demand. Whether countries that have pursued ISI can successfully turn to an open, export-orientated approach will depend to a large extent on the degree of competition they face not only from advanced countries but also from other developing countries.

QUESTIONS

1. How might the interdependence of economies through trade affect an economy's business cycle? Could other forms of interdependence also affect an economy's business cycle?
2. Referring to Table 14.1 on page 421, show how each country could gain from trade if the LDC could produce (before trade) 3 wheat for 1 cloth and the developed country could produce (before trade) 2 wheat for 5 cloth, and if the exchange ratio (with trade) was 1 wheat for 2 cloth. Would they both still gain if the exchange ratio was (a) 1 wheat for 1 cloth and (b) 1 wheat for 3 cloth?
3. Imagine that two countries, Richland and Poorland, can produce just two goods, computers and coal. Assume that for a given amount of land and capital, the output of these two products requires the following constant amounts of labour:

	Richland	Poorland
1 computer	2	4
100 tonnes of coal	4	5

   Assume that each country has 20 million workers.

   a. If there is no trade, and in each country 12 million workers produce computers and 8 million workers produce coal, how many computers and tonnes of coal will each country produce? What will be the total production of each product?
   b. What is the opportunity cost of a computer in (i) Richland; (ii) Poorland?
   c. What is the opportunity cost of 100 tonnes of coal in (i) Richland; (ii) Poorland?
   d. Which country has a comparative advantage in which product?
   e. Assuming that price equals marginal cost, which of the following would represent possible exchange ratios? (i) 1 computer for 40 tonnes of coal; (ii) 2 computers for 140 tonnes of coal; (iii) 1 computer for 100 tonnes of coal; (iv) 1 computer for 60 tonnes of coal; (v) 4 computers for 360 tonnes of coal.
   f. Assume that trade now takes place and that 1 computer exchanges for 65 tonnes of coal. Both countries specialise completely in the product in which they have a comparative advantage. How much does each country produce of its respective product?
   g. The country producing computers sells 6 million domestically. How many does it export to the other country?
   h. How much coal does the other country consume?
4. Why doesn't the USA specialise as much as General Motors or Texaco? Why doesn't the UK specialise as much as Unilever? Is the answer to these questions similar to the answer to the questions, 'Why doesn't the USA specialise as much as Luxembourg?' and 'Why doesn't Unilever specialise as much as the local florist?'
5. To what extent are the arguments for countries specialising and then trading with each other the same as those for individuals specialising in doing the jobs to which they are relatively well suited?
6. The following are four items that are traded internationally: wheat; computers; textiles; insurance. In which one of the four is each of the following most likely to have a comparative advantage: India; the UK; Canada; Japan? Give reasons for your answer.
7. Would it be possible for a country with a comparative disadvantage in a given product at *pre*-trade levels of output to obtain a comparative advantage in it by specialising in its production and exporting it?
8. Go through each of the arguments for restricting trade and provide a counter-argument for not restricting trade.
9. It is often argued that if the market fails to develop infant industries, then this is an argument for government intervention, but not necessarily in the form of restricting imports. What *other* ways could infant industries be given government support?
10. Does the consumer in the importing country gain or lose from dumping? (Consider both the short run and the long run.)
11. What is fallacious about the following two arguments? Is there any truth in either?
    a. 'Imports should be reduced because money is going abroad which would be better spent at home.'
    b. 'We should protect our industries from being undercut by imports produced using cheap labour.'
12. Make out a case for restricting trade between the UK and Japan. Are there any arguments here that could not equally apply to a case for restricting trade between Scotland and England or between Liverpool and Manchester?
13. In what ways may free trade result in harmful environmental effects? Is the best solution to these problems to impose restrictions on trade?
14. If countries are so keen to reduce the barriers to trade, why do many countries frequently attempt to erect barriers?
15. What factors will determine whether a country's joining a customs union will lead to trade creation or trade diversion?
16. How would you set about assessing whether or not a country had made a net dynamic gain by joining a customs union? What sort of evidence would you look for?
17. What would be the economic effects of (a) different rates of VAT; (b) different rates of personal income tax; (c) different rates of company taxation between member states, if in all other respects there were no barriers to trade or factor movements between the members of a customs union?
18. Is trade diversion in the EU more likely or less likely in the following cases?
    a. European producers gain monopoly power in world trade.
    b. Modern developments in technology and communications reduce the differences in production costs associated with different locations.
    c. The development of the internal market produces substantial economies of scale in many industries.
19. Why is it difficult to estimate the magnitude of the benefits of completing the internal market of the EU?

20. Look through the costs and benefits that we identified from the single European market. Do the same costs and benefits arise from the enlarged EU of 27 members as from the 15 members prior to 2004?
21. If there have been clear benefits from the single market programme, why do individual member governments still try to erect barriers, such as new technical standards?
22. Was a 'hard Brexit' (reverting to WTO rules and negotiating bilateral trade deals) necessarily an inferior alternative to remaining in the European single market or, at least, in the customs union?
23. What is meant by the concept of 'learning by exporting'?
24. If a developing country has a comparative advantage in primary products, should the government allow market forces to dictate the pattern of trade?
25. Why will a high exchange rate harm the agricultural sector in a developing country?
26. In what ways may free trade have harmful cultural effects on a developing country?
27. If a developing country has a comparative advantage in the production of wheat, should it specialise as much as possible in the production of wheat and export what is not consumed domestically?
28. What are the advantages and disadvantages for a developing country of pursuing a policy of ISI?
29. Will the production of labour-intensive manufactures for export lead to more or less inequality in a developing country?
30. Would the use of import controls help or hinder a policy of export-orientated industrialisation?
31. Should all developing countries aim over the long term to become *exporters* of manufactured products?

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

14.1 **David Ricardo and the law of comparative advantage.** The first clear statement of the law of comparative advantage (in 1817).
14.2 **Fallacious arguments for restricting trade.** Some of the more common mistaken arguments for protection.
14.3 **Strategic trade theory.** The case of Airbus is used to illustrate the arguments that trade restrictions can be to the strategic advantage of countries in developing certain industries and preventing the establishment of foreign monopolies.
14.4 **The Uruguay Round.** An examination of the negotiations that led to substantial cuts in trade barriers.
14.5 **The World Trade Organization.** This looks at the various opportunities and threats posed by this major international organisation.
14.6 **Steel barriers.** This case considers the US administration's imposition of tariffs on imported steel in 2002, and the reactions of other countries to these tariffs.
14.7 **The Battle of Seattle.** This looks at the protests against the WTO at Seattle in November 1999 and considers the arguments for and against the free trade policies of the WTO.
14.8 **Banana, banana.** The dispute between the USA and the EU over banana imports.
14.9 **Assessing NAFTA.** Who are the winners and losers from NAFTA?
14.10 **Free trade and the environment.** Do whales, the rainforests and the atmosphere gain from free trade?
14.11 **The social dimension of the EU.** The principles of the Social Charter.
14.12 **The benefits of the single market.** Evidence of achievements and the Single Market Action Plan of 1997.
14.13 **The Single Market Scoreboard.** Keeping a tally on progress to a true single market.
14.14 **Multinational corporations and developing economies.** Do MNCs benefit developing economies?
14.15 **Harmonisation in an interdependent world.** Attempts at global policy co-ordination through bodies such as the G7, G20, IMF and WTO.

WEB APPENDICES

14.1 **The gains from trade.** This appendix uses general equilibrium analysis (see Web Appendix 8.1) to analyse the gains from trade and the limits to specialisation under increasing opportunity costs.
14.2 **Exploiting market power in world trade.** A graphical analysis of the effects of a country having monopoly power in the export market or monopsony power in the import market.
14.3 **Trade creation and trade diversion.** A graphical analysis of the welfare gains and losses from joining a customs union.

Balance of payments and exchange rates

In this chapter we will first explain what is meant by the balance of payments. In doing so we will see just how the various monetary transactions between the domestic economy and the rest of the world are recorded.

Then (in Sections 15.2 and 15.3) we will examine how rates of exchange are determined, and how they are related to the balance of payments. We will see what causes exchange rate fluctuations, and how the government can attempt to prevent these fluctuations.

A government could decide to leave its country's exchange rates entirely to market forces (a free-floating exchange rate). Alternatively, it could attempt to fix its currency's exchange rate to some other currency (e.g. the US dollar). Or it could simply try to reduce the degree to which its currency fluctuates. In Section 15.4, we look at the relative merits of different degrees of government intervention in the foreign exchange market: of different 'exchange rate regimes'.

We then turn to look at attempts to achieve greater currency stability between the members of the EU. Section 15.5 looks at the European exchange rate mechanism, which sought in the 1980s and 1990s to limit the amount that member currencies were allowed to fluctuate against each other. Then Section 15.6 examines the euro. Has the adoption of a single currency by 20 EU countries been of benefit to them? To answer this we highlight a series of important factors, such as the proportion of trade between countries within the single currency area.

Finally, as with Chapter 14, we look at the position of developing countries, and this time focus on the issue of debt. Why are so many developing countries facing severe debt problems and what can be done about it?

After studying this chapter, you should be able to answer the following questions:

- What is meant by 'the balance of payments' and how do trade and financial movements affect it?
- How are exchange rates determined and what effects do changes in the exchange rate have on the economy?
- How do governments and/or central banks seek to influence the exchange rate and what are the advantages and disadvantages of such intervention?
- What are the advantages and disadvantages of the euro for member countries? What threats to the stability of the euro arise from the debt and deficit problems of some member states? What can be done to achieve economic growth while tackling these debt problems?
- What are the origins of the severe debt problem faced by many developing countries? What has been done about the problem and what more needs to be done?

15.1 THE BALANCE OF PAYMENTS ACCOUNT

What is meant by a balance of payments deficit or surplus?

In Chapter 9 we identified balance of payments deficits as one of the main macroeconomic problems that governments face. But what precisely do we mean by 'balance of payments deficits' (or surpluses), and what is their significance?

KI 23 p 173

A country's balance of payments account records all the flows of money between residents of that country and the rest of the world. *Receipts* of money from abroad are regarded as credits and are entered in the accounts with a positive sign. *Outflows* of money from the country are regarded as debits and are entered with a negative sign.

There are three main parts of the balance of payments account: the *current account*, the *capital account* and the *financial account*. Each part is then subdivided. We shall look at each part in turn, and take the UK as an example. Table 15.1 gives a summary of the UK balance of payments for 2021, while also providing an historical perspective.

Table 15.1 UK balance of payments

	2021		Average 1987–2021 as % of GDP
	£m	% of GDP	
CURRENT ACCOUNT			
Balance on trade in goods	−156 066	−6.7	−4.5
Balance on trade in services	126 959	5.5	3.2
Balance of trade	**−29 107**	**−1.3**	**−1.3**
Income balance	−11 912	−0.5	−0.4
Net current transfers	−18 931	−0.8	−0.9
Current account balance	**−59 950**	**−2.6**	**−2.6**
CAPITAL ACCOUNT			
Capital account balance	**−2697**	**−0.1**	**0.0**
FINANCIAL ACCOUNT			
Net direct investment	−58 296	−2.5	−0.4
Portfolio investment balance	254 008	11.0	3.2
Other investment balance	−162 370	−7.0	0.1
Balance of financial derivatives	28 647	1.2	0.0
Reserve assets	−17 701	−0.8	−0.2
Financial account balance	**44 288**	**1.9**	**2.6**
Net errors and omissions	**18 359**	**0.8**	**0.0**
Balance	*0.0*	*0.0*	*0.0*

Sources: Based on data from *Balance of Payments time series* and series YBHA (Office for National Statistics)

The current account

The **current account** records payments for imports and exports of goods and services, plus incomes flowing into and out of the country, plus net transfers of money into and out of the country. It is normally divided into four subdivisions.

The trade in goods account. This records imports and exports of physical goods (previously known as 'visibles'). Exports result in an inflow of money and are therefore a credit item. Imports result in an outflow of money and are therefore a debit item. The balance of these is called the **balance on trade in goods** or **balance of visible trade or merchandise balance**. A *surplus* is when exports exceed imports. A deficit is when imports exceed exports.

The trade in services account. This records imports and exports of services (such as transport, tourism and insurance). Thus the purchase of a foreign holiday would be a debit, since it represents an outflow of money, whereas the purchase by an overseas resident of a UK insurance policy would be a credit to the UK services account. The balance of these is called the **services balance**.

The balance of both the goods and services accounts together is known as the **balance on trade in goods and services** or simply the **balance of trade.**

We can relate the balance of trade to our circular flow of income model in Chapter 9. A balance of trade deficit, which as Table 15.1 shows has been the norm in the UK for some time, represents a net leakage from the circular flow because imports (a withdrawal) are greater than exports (an injection). Conversely, a balance of trade surplus is a net injection for an economy.

In equilibrium, injections must equal withdrawals. Thus a net withdrawal on the balance of trade must be

Definitions

Current account of the balance of payments The record of a country's imports and exports of goods and services, plus incomes and transfers of money to and from abroad.

Balance on trade in goods or balance of visible trade or merchandise balance Exports of goods minus imports of goods.

Services balance Exports of services minus imports of services.

Balance on trade in goods and services or balance of trade Exports of goods and services minus imports of goods and services.

offset by a net injection elsewhere: either investment exceeding saving and/or government expenditure exceeding tax revenue. This is why we often see countries with trade deficits running government budget deficits. The USA and the UK are two notable examples.

Countries which run both trade deficits and government budget deficits are said to experience *twin deficits*. A notable exception, however, is Japan which, since the early 1990s, has consistently seen budget deficits accompanied by trade surpluses. Case Study 15.1 on the student website considers further the evidence for twin deficits (and surpluses).

Pause for thought

Can countries run twin surpluses with both a budget surplus and a balance of trade surplus?

Income flows. These consist of wages, interest and profits flowing into and out of the country. For example, dividends earned by a foreign resident from shares in a UK company would be an outflow of money (a debit item).

Current transfers of money. These include government contributions to and receipts from the EU and international organisations, and international transfers of money by private individuals and firms. Transfers out of the country are debits. Transfers into the country (e.g. money sent from Greece to a Greek student studying in the UK) would be a credit item.

The **current account balance** is the overall balance of all the above four subdivisions. A *current account surplus* is where credits exceed debits. A *current account deficit* is where debits exceed credits. Figure 15.1 shows the current account balance expressed as a percentage of GDP for a sample of countries. In conjunction with Table 15.1 we can also see that the UK has consistently run a current account deficit over the past three decades or so. This has been driven by a large trade deficit in goods.

The capital account

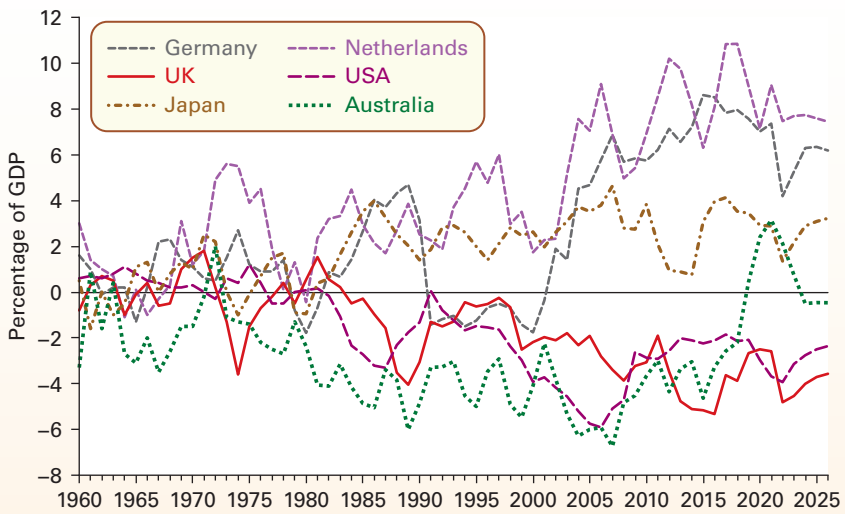
The **capital account** records the flows of funds, into the country (credits) and out of the country (debits), associated with the acquisition or disposal of fixed assets (e.g. land or intangibles, such as patents and trademarks), the transfer of funds by migrants, the payment of grants by the government for overseas projects, debt forgiveness by the government and the receipt of money for capital projects.

Definitions

Balance of payments on current account The balance on trade in goods and services plus net income flows and current transfers.

Capital account of the balance of payments The record of transfers of capital to and from abroad.

Figure 15.1 Current account balance as a percentage of GDP in selected industrial countries



Notes: Figures from 2022 based on forecasts; German figures are West Germany up to 1991

Source: Based on data in *AMECO Database* (European Commission, DGECFIN) (to 1980) and *World Economic Outlook* (IMF, October 2022)

As Table 15.1 shows, the balance on the capital account is small in comparison to that on the current and financial accounts.

The financial account[1]

The **financial account** of the balance of payments records cross-border changes in the holding of shares, property, bank deposits and loans, government securities, etc. In other words, unlike the current account, which is concerned with money incomes, the financial account is concerned with the purchase and sale of assets. Case Study 15.2 on the student website considers some of the statistics behind the UK's financial account.

Direct investment. This involves a significant and lasting interest in a business in another country. If a foreign company invests money from abroad in one of its branches or associated companies in the UK, this represents an inflow of money when the investment is made and is thus a credit item. (Any subsequent profit from this investment that flows abroad will be recorded as an investment income outflow on the current account.) Investment abroad by UK companies represents an outflow of money when the investment is made. It is thus a debit item.

Note that what we are talking about here is the acquisition or sale of assets: e.g. a factory or farm, or the takeover of a whole firm, not the imports or exports of equipment.

Portfolio investment. This relates to transactions in debt and equity securities which do not result in the investor having any significant influence on the operations of a particular business. If a UK resident buys shares (equity securities) in an overseas company, this is an outflow of funds and is hence a debit item.

Other investment and financial flows. These consist primarily of various types of short-term monetary movement between the UK and the rest of the world. Deposits by overseas residents in banks in the UK and loans to the UK from abroad are credit items, since they represent an inflow of money. Deposits by UK residents in overseas banks and loans by UK banks to overseas residents are debit items. They represent an outflow of money.

Short-term monetary flows are common between international financial centres to take advantage of differences in countries' interest rates and changes in exchange rates.

In the financial account, credits and debits are recorded *net*. For example, UK investment abroad consists of the *net* acquisition of assets abroad: i.e. the purchase less the sale of assets abroad. Similarly, foreign investment in the UK consists of the purchase less the sale of UK assets by foreign residents. By recording financial account items net, the flows seem misleadingly modest. For example, if UK residents deposited an extra £100bn in banks abroad but drew out £99bn, this would be recorded as a mere £1bn net outflow on the other investment and financial flows account. In fact, *total* financial account flows vastly exceed current plus capital account flows.

Pause for thought

Where would interest payments on short-term foreign deposits in UK banks be entered on the balance of payments account?

Flows to and from the reserves. The UK, like all other countries, holds reserves of gold and foreign currencies. From time to time the Bank of England (acting as the government's agent) will sell some of these reserves to purchase sterling on the foreign exchange market. It does this normally as a means of supporting the rate of exchange (see below). Drawing on reserves represents a *credit* item in the balance of payments accounts: money drawn from the reserves represents an *inflow* to the balance of payments (albeit an outflow from the reserves account). The reserves can thus be used to support a deficit elsewhere in the balance of payments.

Conversely, if there is a surplus elsewhere in the balance of payments, the Bank of England can use it to build up the reserves. Building up the reserves counts as a debit item in the balance of payments, since it represents an outflow from it (to the reserves).

When all the components of the balance of payments account are taken together, the balance of payments should exactly balance: credits should equal debits. As we shall see below, if they were not equal, the rate of exchange would have to adjust until they were, or the government would have to intervene to make them equal.

When the statistics are compiled, however, a number of errors are likely to occur. As a result there will not be a balance. To 'correct' for this, a **net errors and**

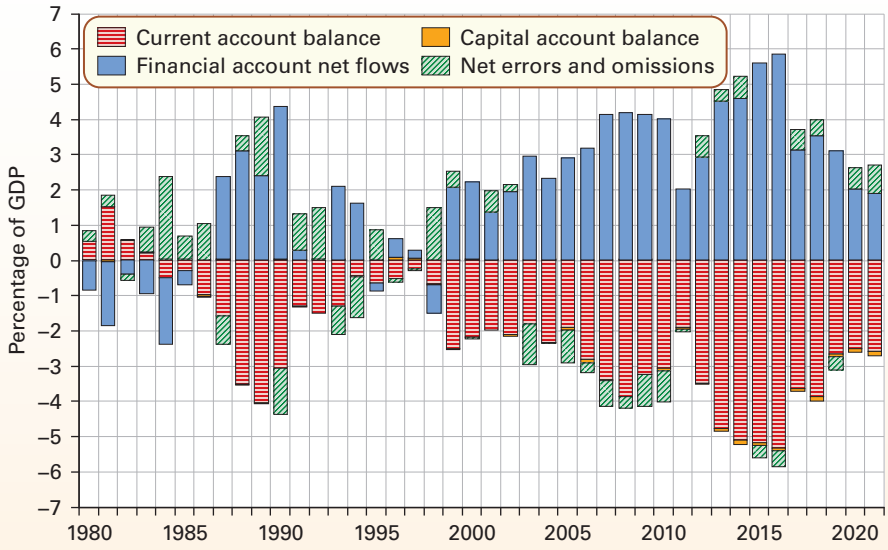
Definitions

Financial account of the balance of payments The record of the flows of money into and out of the country for the purpose of investment or as deposits in banks and other financial institutions.

Net errors and omissions item A statistical adjustment to ensure that the two sides of the balance of payments account balance. It is necessary because of errors in compiling the statistics.

[1]Prior to October 1998 this account was called the 'capital account'. The account that is *now* called the capital account used to be included in the transfers section of the current account. This potentially confusing change of names was adopted in order to bring the UK accounts in line with the system used by the International Monetary Fund (IMF), the EU and most individual countries.

Figure 15.2 UK balance of payments as a percentage of GDP



Source: Based on data from *Balance of Payments time series* and series *YBHA* (Office for National Statistics)

omissions item is included in the accounts. This ensures that there will be an exact balance. The main reason for the errors is that the statistics are obtained from a number of sources, and there are often delays before items are recorded and sometimes omissions too.

Figure 15.2 graphically summarises the main accounts of the UK's balance payments: current, capital and financial accounts. It presents each as a percentage of national income (see also the right-hand column of Table 15.1). In conjunction with the net errors and omissions item, which averages close to zero over the long run, we can see how the accounts combine to give a zero overall balance. For much of the period since the late 1980s, current account deficits have been offset by surpluses on the financial account. The persistence of the UK's current account deficit is discussed further in Case Study 15.3 on the student website.

Recap

1. The balance of payments account records all payments to and receipts from foreign countries.
2. The current account records payments for the imports and exports of goods and services, plus incomes and transfers of money to and from abroad.
3. The capital account records all transfers of capital to and from abroad.
4. The financial account records inflows and outflows of money for investment and as deposits in banks and other financial institutions. These flows represent a transfer of assets rather than income earned. The financial account also includes dealings in the country's foreign exchange reserves.
5. The whole account must balance, but surpluses or deficits can be recorded on any specific part of the account. Thus the current account could be in deficit, but it would have to be matched by an equal and opposite capital plus financial account surplus.

15.2 EXCHANGE RATES

What causes exchange rates to change?

An exchange rate is the rate at which one currency trades for another on the foreign exchange market.

If you want to go abroad, you will need to exchange your pounds into euros, dollars, Swiss francs or whatever. To do this you may go to a bank. The bank will quote you that day's exchange rates: for example, €1.15 to the pound, or $1.25 to the pound. It is similar for firms. If an importer wants to buy, say, some

machinery from Japan, it will require yen to pay the Japanese supplier. It will thus ask the foreign exchange section of a bank to quote it a rate of exchange of the pound into yen. Similarly, if you want to buy some foreign stocks and shares, or if companies based in the UK want to invest abroad, sterling will have to be exchanged into the appropriate foreign currency.

Likewise, if Americans want to come on holiday to the UK or to buy UK assets, or American firms want to import UK goods or to invest in the UK, they will require sterling. They will be quoted an exchange rate for the pound in the USA: say, £1 = $1.25. This means that they will have to pay $1.25 to obtain £1 worth of UK goods or assets.

Exchange rates are quoted between each of the major currencies of the world. These exchange rates are constantly changing. Minute by minute, dealers in the foreign exchange dealing rooms of the banks are adjusting the rates of exchange. They charge commission when they exchange currencies. It is important for them, therefore, to ensure that they are not left with a large amount of any currency unsold. What they need to do is to balance the supply and demand of each currency: to balance the amount they purchase to the amount they sell. To do this they will need to adjust the price of each currency – namely, the exchange rate – in line with changes in supply and demand.

Not only are there day-to-day fluctuations in exchange rates, but also there are long-term changes in them. Figure 15.3 shows the average monthly exchange rates between the pound and various currencies since 1980.

Pause for thought

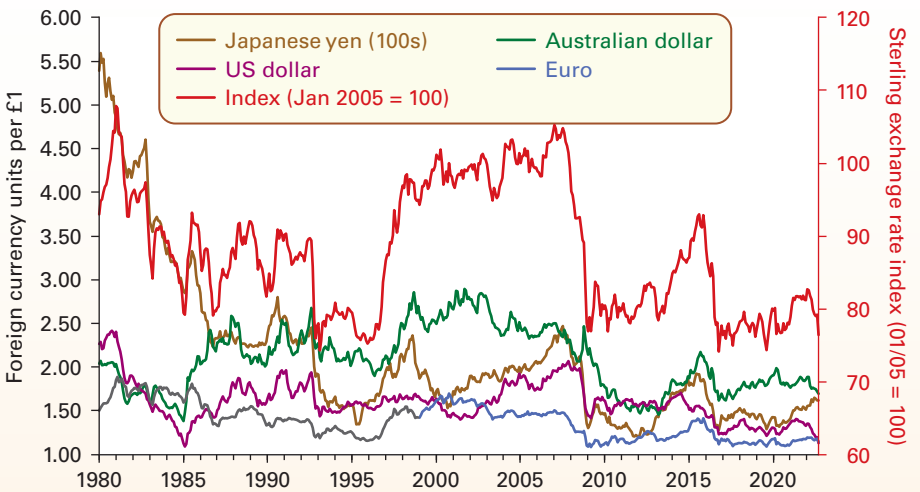
How did the pound 'fare' compared with the US dollar, Australian dollar and the yen from 1980? What conclusions can be drawn about the relative movements of these three currencies?

One of the problems in assessing what is happening to a particular currency is that its rate of exchange may rise against some currencies (weak currencies) and fall against others (strong currencies). In order to gain an overall picture of its fluctuations, therefore, it is best to look at a weighted average exchange rate against all other currencies. This is known as the **exchange rate index** or the **effective exchange rate**. The weight given to each currency in the index depends on the proportion of trade done with that country. Figure 15.3 also shows the sterling exchange rate index based on January 2005 = 100.

Definition

Exchange rate index or effective exchange rate
A weighted average exchange rate expressed as an index, where the value of the index is 100 in a given base year. The weights of the different currencies in the index add up to 1.

Figure 15.3 Sterling exchange rates against selected currencies



Notes: (i) Index is the broad-based effective sterling exchange rate index; (ii) The euro was introduced in 1999, with notes and coins circulating from 2001; (iii) the euro figures prior to 1999 (in grey) are projections backwards in time based on the average exchange rates of the currencies that made up the euro

Source: Based on data in *Statistical Interactive Database* (Bank of England)

BOX 15.1 **NOMINAL AND REAL EXCHANGE RATES** EXPLORING ECONOMICS

Searching for a real advantage

TC 5 p 17

We have seen on several occasions just how important the distinction between nominal and real is. But, what does this distinction mean when applied to exchange rates? A *nominal* bilateral exchange rate is simply the rate at which one currency exchanges for another. All exchange rates that you see quoted in the newspapers, on television or the Internet, or at travel agents, banks or airports, are nominal rates. Up to this point we have solely considered nominal rates.

The *real* exchange rate is the exchange rate index adjusted for changes in the prices of imports (measured in foreign currencies) and exports (measured in domestic prices): in other words, adjusted for the terms of trade. Thus, if a country has a higher rate of inflation for its exports than the weighted average inflation rate of the imports it buys from other countries, its real exchange rate index (RERI) will rise relative to its nominal exchange rate index (NERI).

The real exchange rate index can be defined as:

$$\text{RERI} = \text{NERI} \times P_X/P_M$$

where P_X is the domestic currency price index of exports and P_M is the foreign currencies weighted price index of imports. Thus if (a) a country's inflation rate is 5 per cent higher than the trade-weighted average of its trading partners (P_X/P_M rises by 5 per cent per year) and (b) its nominal exchange rate depreciates by 5 per cent per year (NERI falls by 5 per cent per year), its real exchange rate index will stay the same.

Take another example: if a country's export prices rise faster than the foreign currency prices of its imports (P_X/P_M rises), its real exchange rate will appreciate relative to its nominal exchange rate.

The real exchange rate thus gives us a better idea of the *quantity* of imports a country can obtain from selling a given quantity of exports. If the real exchange rate rises, the country can get more imports for a given volume of exports.

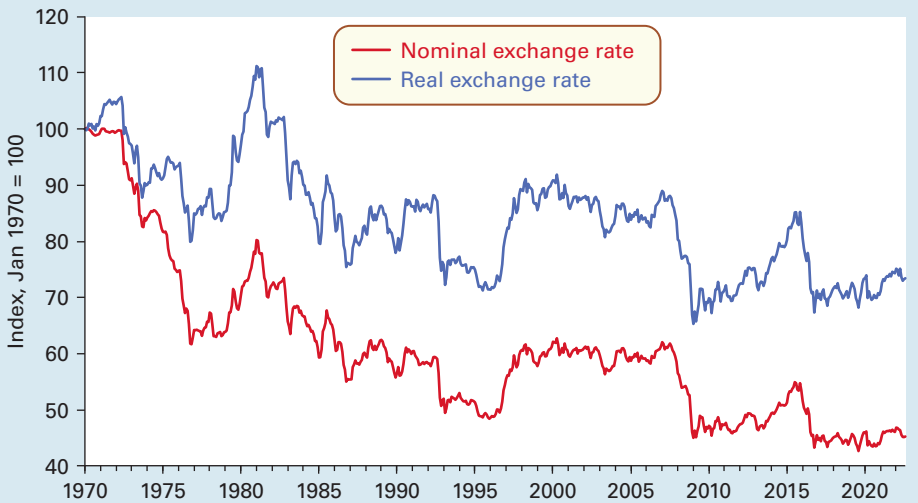
The chart shows the nominal and real exchange rate indices of sterling. As you can see, the real exchange rate has tended to rise over time relative to the nominal exchange rate. This is because the UK has typically had a higher rate of inflation than the weighted average of its trading partners.

The real exchange rate also gives a better idea than the nominal exchange rate of how competitive a country is. The lower the real exchange rate, the more competitive the country's exports will be. From the chart we can see, for example, that the UK became less competitive between 1996 and 2001, and remained at similarly uncompetitive levels until 2008, thanks not only to a rise in the nominal exchange rate index, but also to higher inflation rates than its trading partners. However, with the financial crisis, sterling depreciated sharply. Between July 2007 and October 2009, the nominal and real exchange rate indices both lost almost a quarter of their value. Despite rising in the early 2010s, both indices fell by around 12 per cent in the aftermath of the vote to leave the EU. The indices remained around these levels into the 2020s, but fell temporarily in late 2022 in response to government plans to fund tax cuts (later abandoned) and energy price support measures through substantially increased borrowing.

If differences in inflation rates were to be reflected in longer-term changes in real exchange rates, what pattern should we observe in real exchange rates? Is this supported by the data in the chart?

Download the broad-based effective exchange rate indices from the Bank for International Settlements for the euro, US dollar and British pound. Construct a time-series chart from January 1999 showing the nominal indices for the three currencies and then another chart showing the real indices. Prepare a short PowerPoint presentation containing these charts, which explains the concept of exchange rate indices and the economic significance of the patterns within the charts.

Sterling nominal and real exchange rate indices (Jan 1970 = 100)



Notes: Exchange rate indices are BIS narrow indices comprising 27 countries; re-based by the authors, Jan 1970 = 100

Source: Based on data from *Effective exchange rate indices* (Bank for International Settlements)

Determination of the rate of exchange in a free market

In a free foreign exchange market, the rate of exchange is determined by demand and supply. This is known as a **floating exchange rate** and is illustrated in Figure 15.4.

For simplicity, assume that there are just two countries: the UK and the USA. When UK importers wish to buy goods from the USA, or when UK residents wish to invest in the USA, they *supply* pounds on the foreign exchange market in order to obtain dollars. In other words, they go to banks or other foreign exchange dealers to buy dollars in exchange for pounds. The higher the exchange rate, the more dollars they obtain for their pounds. This effectively makes American goods cheaper to buy, and investment more profitable. Thus the *higher* the exchange rate, the *more* pounds are supplied. The supply curve of pounds therefore typically slopes upwards.

When US residents wish to purchase UK goods or to invest in the UK, they require pounds. They *demand* pounds by selling dollars on the foreign exchange market. In other words, they go to banks or other foreign exchange dealers to buy pounds in exchange for dollars. The lower the dollar price of the pound (the exchange rate), the cheaper it is for them to obtain UK goods and assets, and hence the more pounds they are likely to demand. The demand curve for pounds, therefore, typically slopes downwards.

TC 8 p 37

The equilibrium exchange rate is where the demand for pounds equals the supply. In Figure 15.4 this is at an exchange rate of £1 = $1.30. But what is the mechanism that equates demand and supply?

If the current exchange rate were above the equilibrium, the supply of pounds being offered to the banks would exceed the demand. For example, in Figure 15.4 if the exchange rate were $1.40, there would be an excess supply of pounds of $a - b$. Banks would not have enough dollars to exchange for all these pounds. But the banks make money by *exchanging* currency, not by holding on to it. They would thus lower the exchange rate in order to encourage a greater demand for pounds and reduce the excessive supply. They would continue lowering the rate until demand equalled supply.

Similarly, if the rate were below the equilibrium, say at $1.20, there would be a shortage of pounds of $c - d$. The banks would find themselves with too few pounds to meet all the demand. At the same time they would have an excess supply of dollars. The banks would thus raise the exchange rate until demand equalled supply.

In practice, the process of reaching equilibrium is extremely rapid. The foreign exchange dealers in the banks are continually adjusting the rate as new customers make new demands for currencies. What is more, the banks have to watch closely what each other is doing. They are constantly in competition with each other and thus have to keep their rates in line. The dealers receive

Definition

Floating exchange rate When the government does not intervene in the foreign exchange markets, but simply allows the exchange rate to be freely determined by demand and supply.

Figure 15.4 Determination of the rate of exchange

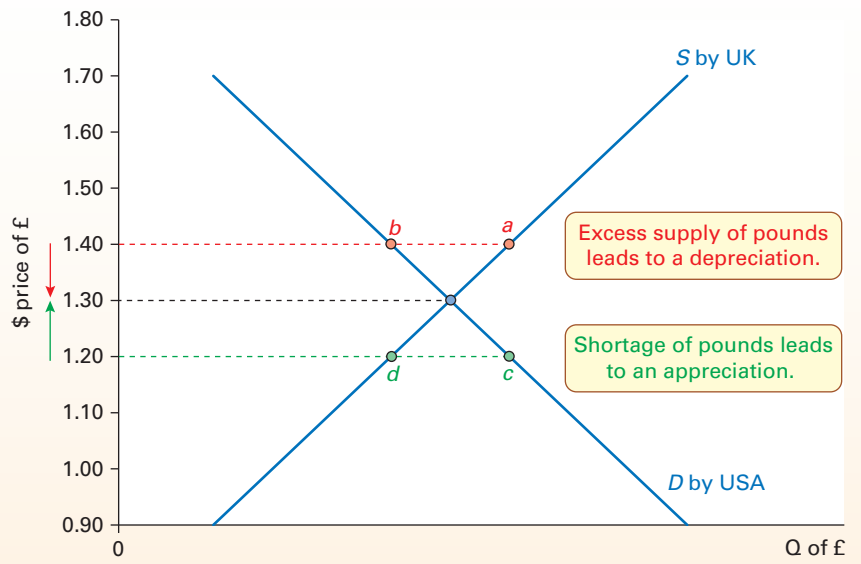
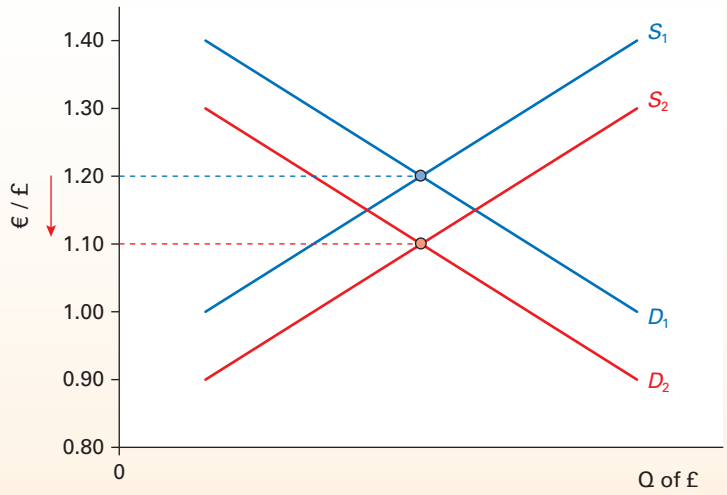


Figure 15.5 Floating exchange rates: movement to a new equilibrium



minute-by-minute updates on their computer screens of the rates being offered around the world.

Shifts in the currency demand and supply curves

KI 8 p 24

Any shift in the currency demand or supply curves will cause the exchange rate to change. This is illustrated in Figure 15.5, but this time by considering the number of euros per £1. If the demand and supply curves shift from D_1 and S_1 to D_2 and S_2 respectively, the exchange rate will fall from €1.20 to €1.10. A fall in the exchange rate is called a **depreciation**. A rise in the exchange rate is called an **appreciation**.

But why should the demand and supply curves shift? The following are the major possible causes of a depreciation:

- *A fall in domestic interest rates*. UK rates would now be less competitive for savers and other depositors. More UK residents would be likely to deposit their money abroad (the supply of sterling would rise), and fewer people abroad would deposit their money in the UK (the demand for sterling would fall).
- *Higher rates of inflation in the domestic economy than abroad*. UK exports will become less competitive. The demand for sterling will fall. At the same time, imports will become relatively cheaper for UK consumers. The supply of sterling will rise.
- *A rise in domestic incomes relative to incomes abroad*. If UK incomes rise, the demand for imports, and hence the supply of sterling, will rise. If incomes in other countries fall, the demand for UK exports, and hence the demand for sterling will fall.
- *Relative investment prospects improving abroad*. If investment prospects become brighter abroad than in the UK, perhaps because of better incentives abroad, or because of worries about an impending recession in the UK, again the demand for sterling will fall and the supply of sterling will rise.

Definitions

Depreciation A fall in the free-market exchange rate of the domestic currency with foreign currencies.

Appreciation A rise in the free-market exchange rate of the domestic currency with foreign currencies.

Recap

1. The rate of exchange is the rate at which one currency exchanges for another. Rates of exchange are determined by demand and supply in the foreign exchange market. Demand for the domestic currency consists of all the credit items in the balance of payments account. Supply consists of all the debit items.
2. The exchange rate will depreciate (fall) if the demand for the domestic currency falls or the supply increases. These shifts can be caused by a fall in the domestic interest rates, higher inflation in the domestic economy than abroad, a rise in domestic incomes relative to incomes abroad, relative investment prospects improving abroad or the belief by speculators that the exchange rate will fall.
3. The opposite in each case would cause an appreciation (rise).

- *Speculation that the exchange rate will fall.* If businesses involved in importing and exporting, and also banks and other foreign exchange dealers, think that the exchange rate is about to fall, they will sell pounds now before the rate does fall. The supply of sterling will thus rise.

Pause for thought

Go through each of the above reasons for shifts in the demand for and supply of sterling and consider what would cause an appreciation of the pound.

15.3 EXCHANGE RATES AND THE BALANCE OF PAYMENTS

How does the balance of payments affect the exchange rate?

Exchange rates and the balance of payments: no government or central bank intervention

KI 8 p 24

In a free foreign exchange market, the balance of payments will *automatically* balance. But why?

The credit side of the balance of payments constitutes the demand for sterling. For example, when people abroad buy UK exports or assets they demand sterling in order to pay for them. The debit side constitutes the supply of sterling. For example, when UK residents buy foreign goods or assets, the importers of them require foreign currency to pay for them. They thus supply pounds. A floating exchange rate ensures that the demand for pounds is equal to the supply. It thus also ensures that the credits on the balance of payments are equal to the debits: that the balance of payments balances.

This does not mean that each part of the balance of payments account separately balances, but simply that any current account deficit must be matched by a capital plus financial account surplus and vice versa.

For example, suppose initially that each part of the balance of payments *did* separately balance. Then let us assume that interest rates rise. This encourages larger short-term financial inflows as people abroad are attracted to deposit money in the UK: the demand for sterling would shift to the right (e.g. from D_2 to D_1 in Figure 15.5). It will also cause smaller short-term financial outflows as UK residents keep more of their money in the country: the supply of sterling shifts to the left (e.g. from S_2 to S_1 in Figure 15.5). The financial account will go into surplus. The exchange rate will appreciate.

As the exchange rate rises, this will cause imports to be cheaper and exports to be more expensive.

BOX 15.2 DEALING IN FOREIGN CURRENCIES — CASE STUDIES & APPLICATIONS

A daily juggling act

Imagine that a large car importer in the UK wants to import 5000 cars from Japan costing ¥15 billion. What does it do?

It will probably contact a number of banks' foreign exchange dealing rooms in London and ask them for exchange rate quotes. It thus puts all the banks in competition with each other. Each bank will want to get the business and thereby obtain the commission on the deal. To do this it must offer a higher rate than the other banks, since the higher the ¥/£ exchange rate, the more yen the firm will get for its money. (For an importer a rate of, say, ¥160 to £1 is better than a rate of, say, ¥140.)

Now it is highly unlikely that any of the banks will have a spare ¥15 billion. But a bank cannot say to the importer 'Sorry, you will have to wait before we can agree to sell them to you.' Instead the bank will offer a deal and then, if the firm agrees, the bank will have to set about obtaining the ¥15 billion. To do this it must offer the Japanese who are supplying yen the chance to obtain pounds at a sufficiently *low* ¥/£ exchange rate. (The lower the ¥/£ exchange rate, the fewer yen the Japanese will have to pay to obtain pounds.)

The banks' dealers thus find themselves in the delicate position of wanting to offer a *high* enough exchange rate to the car importer in order to gain its business, but a *low* enough exchange rate in order to obtain the required amount of yen. The dealers are thus constantly having to adjust the rates of exchange in order to balance the demand and supply of each currency.

In general, the more of any foreign currency that dealers are asked to supply (by being offered sterling), the lower the exchange rate they will offer will be. In other words, a higher supply of sterling pushes down the foreign currency price of sterling.

Assume that an American firm wants to import Scotch whisky from the UK. Describe how foreign exchange dealers will respond.

Download the latest edition of the Triennial Central Bank Survey of foreign exchange markets published by the Bank for International Settlements (BIS). Briefly summarise the levels of activity on the foreign exchange markets, including details of the principal currencies being bought and sold.

TC 8 p 37 The current account will move into deficit. There is a movement up along the new demand and supply curves until a new equilibrium is reached. At this point, any financial account surplus is matched by an equal current (plus capital) account deficit.

Exchange rates and the balance of payments: with government or central bank intervention

TC 11 p 61 The government or central bank may be unwilling to let the country's currency float freely. Frequent shifts in the demand and supply curves would cause frequent changes in the exchange rate. This, in turn, might cause uncertainty for businesses, which might curtail their trade and investment.

The central bank may thus intervene in the foreign exchange market. But what can it do? The answer to this depends on its objectives. It may simply want to reduce the day-to-day fluctuations in the exchange rate, or it may want to prevent longer-term, more fundamental shifts in the rate.

Reducing short-term fluctuations

Assume that the UK government believes that an exchange rate of €1.20 to the pound is approximately the long-term equilibrium rate. Short-term leftward shifts in the demand for sterling and rightward shifts in the supply, however, are causing the exchange rate to fall below this level (see Figure 15.5). What can the government do to keep the rate at €1.20?

Using reserves. The Bank of England can sell gold and foreign currencies from the reserves to buy pounds. This will shift the demand for sterling back to the right. However, with the growth of short-term international financial flows it is in practice very difficult for individual central banks to influence exchange rates significantly by buying and selling currencies. The combined actions of central banks might, however, be more successful.

Borrowing from abroad. In extreme circumstances, the government could negotiate a foreign currency loan from other countries or from an international agency such as the International Monetary Fund. It can then use these moneys to buy pounds on the foreign exchange market, thus again shifting the demand for sterling back to the right.

Raising interest rates. If the Bank of England raises interest rates, it will encourage people to deposit money in the UK and encourage UK residents to keep their money in the country. The demand for sterling will increase and the supply of sterling will decrease. However, the changes in interest rates necessary to manage the exchange rate may come into conflict with other economic objectives, such as keeping the rate of inflation on target.

Maintaining a fixed rate of exchange over the longer term

Governments may choose to maintain a fixed rate over a number of months or even years. The following are possible methods it can use to achieve this (we are assuming that there are downward pressures on the exchange rate: e.g. as a result of higher aggregate demand and higher inflation).

Contractionary policies. This is where the government deliberately curtails aggregate demand by either *fiscal policy* or *monetary policy* or both.

Contractionary fiscal policy involves raising taxes and/or reducing government expenditure. Contractionary monetary policy involves raising interest rates. Note that in this case we are talking about not just the temporary raising of interest rates to prevent a short-term outflow of money from the country, but the use of higher interest rates to reduce borrowing and hence dampen aggregate demand.

A reduction in aggregate demand works in two ways:

- It reduces the level of consumer spending. This directly cuts imports, since there will be reduced spending on Japanese electronics, German cars, Spanish holidays and so on. The supply of sterling coming on to the foreign exchange market thus decreases.
- It reduces the rate of inflation. If inflation falls below that of other countries, this makes UK goods more competitive abroad, thus increasing the demand for sterling. It will also cut back on imports as UK consumers switch to the now more competitive home-produced goods. The supply of sterling falls.

Supply-side policies. This is where the government attempts to increase the long-term competitiveness of UK goods by encouraging reductions in the costs of production and/or improvements in the quality of UK goods. For example, the government may attempt to improve the quantity and quality of training and research and development (see Section 13.5).

Controls on imports and/or foreign exchange dealing. This is where the government restricts the outflow of money, either by restricting people's access to foreign exchange, or by the use of tariffs (customs duties) and quotas. For instance, the Icelandic government put in place controls on foreign currency exchanges in the aftermath of the collapse of its largest banks in 2008 in an attempt to bolster the krona and to build up foreign reserves.

Pause for thought

What problems might arise if the government were to adopt this third method of maintaining a fixed exchange rate?

Recap

1. In a free foreign exchange market, the balance of payments will automatically balance, since changes in the exchange rate will balance the demand for the currency (credits on the balance of payments) with the supply (debits on the balance of payments).
2. There is no guarantee, however, that there will be a balance on each of the separate parts of the balance of payments account.
3. The government can attempt to prevent the rate of exchange falling by central bank purchases of the domestic currency in the foreign exchange market, either by selling foreign currency reserves or by using foreign loans. Alternatively, the government can raise interest rates. The reverse actions can be taken if it wants to prevent the rate from rising.
4. In the longer term, it can attempt to prevent the rate from falling by pursuing contractionary policies, protectionist policies or supply-side policies to increase the competitiveness of the country's exports.

15.4 FIXED VERSUS FLOATING EXCHANGE RATES

Should exchange rates be 'left to the market'?

Are exchange rates best left free to fluctuate and be determined purely by market forces, or should the government or central bank intervene to fix exchange rates, either rigidly or within bands? Unfortunately, the answer is not clear-cut. Both floating and fixed exchange rates have their advantages and disadvantages.

Advantages of fixed exchange rates

Many economists are opposed to fixed exchange rates, for reasons to be examined shortly. Nevertheless, many businesspeople are in favour of relatively rigid exchange rates. The following arguments are used.

KI 15 p 80

Certainty. With fixed exchange rates, international trade and investment become much less risky, since profits are not affected by movements in the exchange rate.

Assume that a firm correctly forecasts that its product will sell in the USA for $1.50. It costs 80p to produce. If the rate of exchange is fixed at £1 = $1.50, each unit will earn £1 and hence make a 20p profit. If, however, the rate of exchange were not fixed, exchange fluctuations could wipe out this profit. If, say, the rate appreciated to £1 = $2, and if units continued to sell for $1.50, they would now earn only 75p each, and hence make a 5p loss.

Little or no speculation. Provided the rate is *absolutely* fixed – and people believe that it will remain so – there is no point in speculating. For example, between 1999 and 2001, when the old currencies of the eurozone countries were still used, but were totally fixed to the euro, there was no speculation that the German mark, say, would change in value against the French franc or the Dutch guilder.

Prevents governments pursuing 'irresponsible' macroeconomic policies. If a government deliberately and excessively expands aggregate demand – perhaps in an attempt to gain short-term popularity with the electorate – the resulting balance of payments deficit will force it to constrain demand again (unless it resorts to import controls).

Governments cannot allow their economies to have a persistently higher inflation rate than competitor countries without running into balance of payments crises, and hence a depletion of reserves. Fixed rates thus force governments (in the absence of trade restrictions) to keep the domestic rate of inflation close to the world rate.

Disadvantages of fixed exchange rates

KI 31 p 206

Exchange rate policy may conflict with the interests of domestic business and the economy as a whole. A balance of payments deficit can occur even if the economy is not 'overheating'. For example, there can be a fall in the demand for the country's exports as a result of an external shock (such as a recession in other countries) or because of increased foreign competition. If protectionism is to be avoided, and if supply-side policies work only over the long run, the government will be forced to raise interest rates. This is likely to have two adverse effects on the domestic economy:

- Higher interest rates may discourage long-term business investment. This in turn will lower firms' profits in the long term and reduce the country's long-term rate of economic growth. The country's capacity to produce will be restricted and businesses are likely to fall behind in the competitive race with their international rivals to develop new products and improve existing ones.

TC 15 p 287

- Higher interest rates will have a dampening effect on the economy by making borrowing more expensive and thereby cutting back on both consumer demand and investment. This can result in a recession with rising unemployment.

The problem is that, with fixed exchange rates, domestic policy is entirely constrained by the balance of payments. Any attempt to reflate and cure unemployment will simply lead to a balance of payments deficit and thus force governments to deflate again.

Competitive contractionary policies leading to world recession. TC 14 p 249 If deficit countries pursued contractionary policies, but surplus countries pursued expansionary policies, there would be no overall world contraction or expansion. Countries may be quite happy, however, to run a balance of payments surplus and build up reserves. Countries may thus competitively deflate – all trying to achieve a balance of payments surplus. But this is beggar-my-neighbour policy. Not all countries can have a surplus! Overall the world must be in balance. The result of these policies is to lead to a general world recession and a restriction in growth.

Problems of international liquidity. If trade is to expand, there must be an expansion in the supply of currencies acceptable for world trade (dollars, euros, gold, etc.): there must be adequate **international liquidity**. Countries' reserves of these currencies must grow if they are to be sufficient to maintain a fixed rate at times of balance of payments disequilibrium. Conversely, there must not be excessive international liquidity; otherwise the extra demand that would result would lead to world inflation. It is important under fixed exchange rates, therefore, to avoid too much or too little international liquidity.

The problem is how to maintain adequate control of international liquidity. The supply of dollars, for example, depends largely on US policy, which may be dominated by its internal economic situation rather than by any concern for the well-being of the international community. Similarly, the supply of euros depends on the policy of the European Central Bank, which is governed by the internal situation in the eurozone countries.

TC 3 p 10 *Inability to adjust to shocks.* With sticky prices and wage rates, there is no swift mechanism for dealing with sudden balance of payments crises – like that caused by a sudden increase in oil prices. In the short run, countries will need huge reserves or loan facilities to support their currencies. There may be insufficient international liquidity to permit this. In the longer run, countries may be forced into a depression by having to deflate. The alternative may be to resort to protectionism, or to abandon the fixed rate and **devalue**.

TC 11 p 61 *Speculation.* If speculators believe that a fixed rate simply cannot be maintained, speculation is likely to be massive. If, for example, there is a large balance of payments deficit, speculative selling will worsen the deficit and may itself force a devaluation. Speculation of this sort had disastrous effects on some south-east Asian currencies in 1997 (see Case Study 15.9 on the student website) and on the Argentinean peso in 2002 (see Case Study 15.10).

Advantages of a free-floating exchange rate

The advantages and disadvantages of free-floating rates are to a large extent the opposite of fixed rates.

Automatic correction. The government simply lets the exchange rate move freely to the equilibrium. In this way, balance of payments disequilibria are automatically and instantaneously corrected without the need for specific government policies. KI 8 p 24

No problem of international liquidity and reserves. Since there is no central bank intervention in the foreign exchange market, there is no need to hold reserves. A currency is automatically convertible at the current market exchange rate.

Insulation from external economic events. A country is not tied to a possibly unacceptably high world inflation rate, as it could be under a fixed exchange rate. It is also to some extent protected against world economic fluctuations and shocks. TC 3 p 10

Governments are free to choose their domestic policy. Under a floating rate, the government can choose whatever level of domestic demand it considers appropriate and simply leave exchange rate movements to take care of any balance of payments effect. Similarly, the central bank can choose whatever rate of interest is necessary to meet domestic objectives, such as achieving a target rate of inflation. The exchange rate will simply adjust to the new rate of interest – a rise in interest rates causing an appreciation, a fall causing a depreciation.

This freedom for the government and central bank is a major advantage, especially when the effectiveness of contractionary policies under fixed exchange rates is reduced by downward wage and price rigidity, and when competitive contractionary policies between countries may end up causing a world recession.

Definitions

International liquidity The supply of currencies in the world acceptable for financing international trade and investment.

Devaluation Where the government refixes the exchange rate at a lower level.

BOX 15.3 THE IMPORTANCE OF INTERNATIONAL FINANCIAL MOVEMENTS EXPLORING ECONOMICS

How a current account deficit can coincide with an appreciating exchange rate

Since the early 1970s most of the major economies of the world have operated with floating exchange rates. The opportunities that this gives for speculative gain has led to a huge increase in short-term international financial movements. Vast amounts of moneys transfer from country to country in search of higher interest rates or a currency that is likely to appreciate. This can have a bizarre effect on exchange rates.

If a country pursues an expansionary fiscal policy, the current account will tend to go into deficit as extra imports are 'sucked in'. What effect will this have on exchange rates? You might think that the answer is obvious: the higher demand for imports will create an extra supply of domestic currency on the foreign exchange market and hence drive down the exchange rate.

In fact, the opposite is likely. The higher interest rates resulting from the higher domestic demand can lead to a massive inflow of short-term finance. The financial account can thus move sharply into surplus. This is likely to outweigh the current account deficit and cause an *appreciation* of the exchange rate.

Exchange rate movements, especially in the short term, are largely brought about by changes on the financial rather than the current account.

Why do high international financial mobility and an absence of exchange controls severely limit a country's ability to choose its interest rate?

Undertake a literature search on the topic of interest rates and exchange rate movements. Write a short review of your findings.

Disadvantages of a free-floating exchange rate

Despite these advantages there are still some potentially serious problems with free-floating exchange rates.

TC 10 p 49 *Unstable exchange rates.* The less elastic the demand and supply curves are for the currency in Figure 15.5, the greater the change in exchange rate that will be necessary to restore equilibrium following a shift in either demand or supply will be. In the long run, in a competitive world with domestic substitutes for imports and foreign substitutes for exports, demand and supply curves are relatively elastic. Nevertheless, in the short run, given that many firms have contracts with specific overseas suppliers or distributors, the demands for imports and exports are less elastic.

TC 11 p 61 *Speculation.* Short-run instability can be lessened by stabilising speculation, thus making speculation advantageous. If, due to short-run inelasticity of demand, a deficit causes a very large depreciation, speculators will *buy* pounds, knowing that in the long run the exchange rate will appreciate again. Their action therefore helps to lessen the short-run fall in the exchange rate.

Nevertheless, in an uncertain world where there are few restrictions on currency speculation, where the fortunes and policies of governments can change rapidly, and where large amounts of short-term deposits are internationally 'footloose', speculation can be highly destabilising in the short run. Considerable exchange rate overshooting can occur.

An example of such overshooting occurred between July 2008 and March 2009 when the pound depreciated 14 per cent against the euro, 29 per cent against the US dollar and 35 per cent against the yen (see Figure 15.3 on page 458). The nominal sterling exchange rate index fell by 17 per cent (see Figure in Box 15.1). Speculators were predicting that interest rates in the UK would fall further than in other countries and stay lower for longer. This was because the recessionary implications of a highly-financialised economy such as the UK following the financial crisis were expected to be deeper. Consequently, inflation was expected to undershoot the Bank of England's 2 per cent target and perhaps even become negative. But the fall in the exchange rate represented considerable overshooting and the nominal exchange rate index rose 8.5 per cent between March and June 2009.

This is just one example of the violent swings in exchange rates that have occurred in recent years. They even occur under managed floating exchange rate systems where governments have attempted to dampen such fluctuations!

The continuance of exchange rate fluctuations over a number of years is likely to encourage the growth of speculative holdings of currency. This can then cause even larger and more rapid swings in exchange rates.

Pause for thought

If speculators on average gain from their speculation, who loses?

Uncertainty for traders and investors. The uncertainty caused by currency fluctuations can discourage international trade and investment. To some extent

KI 15 p 80

the problem can be overcome by using the **forward exchange market**. Here traders agree with a bank *today* the rate of exchange for some point in the future (say, six months' time). This allows traders to plan future purchases of imports or sales of exports at a known rate of exchange. Of course, banks charge for this service, since they are taking on the risks themselves of adverse exchange rate fluctuations.

But dealing in the futures market only takes care of short-run uncertainty. Banks will not be prepared to take on the risks of offering forward contracts for several years hence. Thus firms simply have to live with the uncertainty over exchange rates in future years. This may discourage long-term investment. For example, the possibility of exchange rate appreciation may well discourage firms from investing abroad, since a higher exchange rate means that foreign exchange earnings will be worth less in the domestic currency.

As Figure 15.3 showed (see page 458) there have been large changes in exchange rates. Such changes make it difficult not only for exporters. Importers too will be hesitant about making long-term deals. For example, a UK manufacturing firm signing a contract to buy US components in March 2008, when $2.00 worth of components could be purchased for £1, would find it a struggle to make a profit some five years later when less than $1.50 worth of US components could be purchased for £1!

Lack of discipline on the domestic economy. Governments may pursue irresponsibly inflationary policies (for short-term political gain, say). This will have adverse effects over the longer term as the government will at some point have to deflate the economy again, with a resulting fall in output and rise in unemployment.

Exchange rates in practice

Exchange rates have become extremely volatile. Currencies can gain or lose several percentage points in the space of a few days. These changes can then make all the difference between profit and loss for trading companies. There are a number of reasons for this volatility:

- Inflation or money supply targets. Central banks may have to make considerable changes to interest rates in order to keep to their targets. These in turn cause exchange rate fluctuations.
- A huge growth in international financial markets. The Bank for International Settlements estimated that in 2019 trading in foreign exchange markets had reached $6.6 trillion per day.
- The abolition of exchange controls in most industrialised countries.
- The growth in information technology. The simple use of a computer can transfer capital and finance internationally in a matter of seconds.
- The preference for liquidity. With the danger of currency fluctuations, companies prefer to keep their financial capital as liquid as possible. They do not want to be locked into assets denominated in a declining currency.
- The growing speculative activities of trading companies. Many large companies have a team of dealers to help manage their liquid assets: to switch them from currency to currency in order to take advantage of market movements.

Definition

Forward exchange market Where contracts are made today for the price at which a currency will be exchanged at some specified future date.

Recap

1. Completely fixed exchange rates bring the advantage of certainty for the business community, which encourages trade and foreign investment. They also help to prevent governments from pursuing irresponsible macroeconomic policies.
2. However, with fixed rates domestic policy is entirely constrained by the balance of payments. What is more, they can lead to competitive contractionary policies worldwide; there may be problems of excessive or insufficient international liquidity; there may be difficulty in adjusting to external shocks; and speculation could be very severe if people came to believe that a fixed rate was about to break down.
3. The advantages of free-floating exchange rates are that they automatically correct balance of payments disequilibria; they eliminate the need for reserves; and they give governments a greater independence to pursue their chosen domestic policy.
4. On the other hand, a completely free exchange rate can be highly unstable, especially when the elasticities of demand for imports and exports are low and there are shifts in currency demand and supply; in addition, speculation may be destabilising. This may discourage firms from trading and investing abroad. What is more, a flexible exchange rate, by removing the balance of payments constraint on domestic policy, may encourage governments to pursue irresponsible domestic policies for short-term political gain.

BOX 15.4 THE EURO/DOLLAR SEE-SAW

Ups and downs in the currency market

For periods of time, world currency markets can be quite peaceful, with only modest changes in exchange rates. But with the ability to move vast sums of money very rapidly from one part of the world to another and from one currency to another, speculators can suddenly turn this relatively peaceful world into one of extreme turmoil.

In this box we examine the huge swings of the euro against the dollar since the euro's launch in 1999.

First the down. . .

On 1 January 1999, the euro was launched and exchanged for $1.16. By October 2000 the euro had fallen to $0.85. The main cause of this 27 per cent depreciation was the growing fear that inflationary pressures were increasing in the USA and that, therefore, the Federal Reserve Bank would have to raise interest rates. At the same time, the eurozone economy was growing only slowly and inflation was well below the 2 per cent ceiling set by the ECB. There was thus pressure on the ECB to cut interest rates.

The speculators were not wrong. As the diagram shows, US interest rates rose, and ECB interest rates initially fell, and when eventually they did rise (in October 1999), the gap between US and ECB interest rates soon widened again.

In addition to the differences in interest rates, a lack of confidence in the recovery of the eurozone economy and a continuing confidence in the US economy encouraged investment to flow to the USA. This inflow of finance (and lack of inflow to the eurozone) further pushed up the dollar relative to the euro.

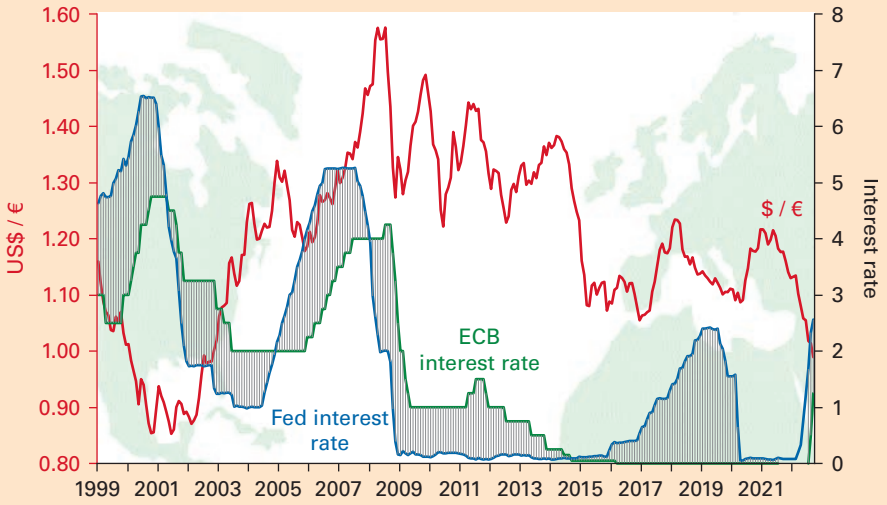
The low value of the euro against the dollar meant a high value of other currencies, including the pound, relative to the euro. This made it very difficult for companies outside the eurozone to export to eurozone countries and also for those competing with imports from the eurozone (which had been made cheaper by the fall in the euro).

In October 2000, with the euro trading at around 85¢, the ECB plus the US Federal Reserve Bank, the Bank of England and the Japanese central bank all intervened on the foreign exchange market to buy euros. This arrested the fall, and helped to restore confidence in the currency.

. . . then the up

The position changed completely in 2001. With the US economy slowing rapidly and fears of an impending recession, the Federal Reserve Bank reduced interest rates 11 times during the year: from 6.5 per cent at the beginning of the year to 1.75 per cent at the end (see the diagram). Although the ECB also cut interest rates, the cuts were relatively modest: from 4.75 at the beginning of the year to 3.25 at the end. With eurozone interest rates now considerably above US rates, the euro began to rise.

Fluctuations between the euro and the dollar



Notes: Federal reserve rate is the federal funds *effective* rate (monthly average); ECB interest rate is the main refinancing operations rate (end of month)

Sources: $/€ based on data from *Statistical Interactive Database*, Bank of England, series XUMAERD (data published 3 October 2022); interest rate data from *Federal Reserve Bank* and *European Central Bank*

- The growing speculative activities of banks and other financial institutions.
- The growing belief that rumour and 'jumping on the bandwagon' are more important determinants of currency buying or selling than cool long-term appraisal. If people believe that speculation is likely to be destabilising, their actions will ensure that it is. Many companies involved in

CASE STUDIES & APPLICATIONS

In addition, a massive deficit on the US current account, and a budget deficit nearing 4 per cent of GDP, made foreign investors reluctant to invest in the US economy. In fact, investors were pulling out of the USA. One estimate suggests that European investors alone sold $70 billion of US assets during 2002. The result of all this was a massive depreciation of the dollar and appreciation of the euro, so that by December 2004 the euro had risen to $1.36: a 60 per cent appreciation since July 2001!

In 2004–5 with the US economy growing strongly again, the Fed raised interest rates several times, from 1 per cent in early 2004 to 5.25 by June 2006. With growth in the eurozone averaging just 1.8 per cent in 2004–5, the ECB kept interest rates constant at 2 per cent until early 2006. The result was that the euro depreciated against the dollar in 2005. But then the rise of the euro began again as the US growth slowed and eurozone growth rose and people anticipated a narrowing of the gap between US and eurozone interest rates.

In 2007 and 2008 worries about the credit crunch in the USA led the Fed to make substantial cuts in interest rates to stave off recession. In August 2007 the US federal funds rate was 5.25 per cent. It was then reduced on several occasions to stand at between 0 and 0.25 per cent by December 2008. The ECB, in contrast, kept the eurozone rate constant at 4 per cent for the first part of this period and even raised it to 4.25 temporarily in the face of rapidly rising commodity prices. As a result, short-term finance flooded into the eurozone and the euro appreciated again, from $1.37 in mid-2007 to $1.58 in mid-2008.

. . . then the steps down

Eventually, in September 2008, with the eurozone on the edge of recession and predictions that the ECB would cut interest rates, the euro at last began to fall. It continued to do so as the ECB cut rates. However, with monetary policy in the eurozone remaining tighter than in the USA, the euro began to rise again, only falling once more at the end of 2009 and into 2010 as US growth accelerated and speculators anticipated a tightening of US monetary policy.

Growing worries in 2010 about the level of government deficits and debt in various eurozone countries, such as Greece, Portugal, Spain, Italy and Ireland, contributed to speculation and thus growing volatility of the euro. Throughout the first part of 2010 investors became increasingly reluctant to hold the euro as fears of debt default mounted. As such, the euro fell substantially from $1.44 in January 2010 to $1.19 in June. This was a 17 per cent depreciation.

Then, as support was promised by the ECB and IMF to Greece in return for deficit reduction policies, and similar support could be made available to other eurozone countries with severe deficits, fears subsided and the euro rose again. By the end of October 2010, the euro was trading at $1.39. In April 2011, the euro rose to a high of $1.44.

Then began a dramatic fall in the euro as concerns grew over the eurozone's sluggish recovery and continuing high debt levels. Speculators thus believed that eurozone interest rates would have to continue falling. The ECB cut the main interest rate from 1.5 per cent in October 2011 in a series of steps to 0.05 per cent by September 2014.

With the ECB reducing interest rates and people increasingly predicting the introduction of quantitative easing (QE), the euro depreciated during 2014. Between March and December 2014 it depreciated by 11 per cent against the dollar, while the euro exchange rate index depreciated by 4 per cent. With the announced programme of QE being somewhat larger than markets expected, in the week following the announcement in January 2015, the euro fell a further 2.3 per cent against the dollar, and the euro exchange rate index also fell by 2.3 per cent. The result was that the euro was trading at its lowest level against the US dollar since April 2003.

The euro strengthened against the dollar in 2017/18 as economic growth in the eurozone picked up. But this economic respite was to prove only temporary with the growth rate falling to only 1.5 per cent in 2019. After pausing in December 2018 (see Box 22.7), quantitative easing resumed in November 2019. The euro was on the slide again.

The euro-dollar see-saw continued during the early 2020s. As the COVID-19 pandemic took hold, investors saw the dollar as a 'haven of last resort' and the dollar strengthened. It eased as the USA struggled to control the spread of COVID cases and thus its impact on economic activity. Then, as supply-side disruptions, exacerbated by the Russian invasion of Ukraine, fuelled inflation, US monetary policy tightened more quickly and aggressively than in the eurozone. The dollar therefore strengthened again and by September 2022 the euro was worth less than a dollar.

The path of the euro shows that the monetary policy decisions of the ECB and the Fed have been a major factor in the exchange-rate volatility between the euro and the dollar – itself a cause of uncertainty in international trade and finance. However, other factors such as the wellbeing of countries' public finances, controlling the spread of COVID-19 and global political tensions have contributed to fluctuations in the euro-dollar see-saw.

How important are relative interest rates in the long run in the determination of bilateral exchange rates, such as that between the dollar and the euro?

Find out what has happened to the euro/dollar exchange rate over the past 12 months. You can find the data from the Bank of England's Statistical Interactive Database. Explain why the exchange rate has moved the way it has.

international trade and finance have developed a 'speculative mentality'.

- The growing belief that governments are powerless to prevent currency movements. As short-term capital (or 'hot money') grows relative to official reserves, it is increasingly difficult for central banks to stabilise currencies through exchange market intervention.

15.5 THE ORIGINS OF THE EURO

How did the majority of the EU countries arrive at a single currency?

There have been many attempts to regulate exchange rates since 1945. By far the most successful was the Bretton Woods system, which was adopted worldwide from the end of the Second World War until 1971. This was a form of **adjustable peg** exchange rate, where countries pegged (i.e. fixed) their exchange rate to the US dollar, but could re-peg it at a lower or higher level ('devalue' or 'revalue' their exchange rate) if there was a persistent and substantial balance of payments deficit or surplus.

TC 11 p 61

With growing world inflation and instability from the mid-1960s, it became more and more difficult to maintain fixed exchange rates, and the growing likelihood of devaluations and revaluations fuelled speculation. The system was abandoned in the early 1970s. What followed was a period of exchange rate management known as **managed floating**. Under this system, exchange rates were not pegged but allowed to float. However, central banks intervened from time to time to prevent excessive exchange rate fluctuations. This system largely continues to this day.

However, on a regional basis, especially within Europe, there were attempts to create greater exchange rate stability. The European system involved establishing exchange rate bands: upper and lower limits within which exchange rates were allowed to fluctuate. The name given to the EU system was the **exchange rate mechanism** (**ERM**).

The ERM

The ERM came into existence in March 1979 and the majority of the EU countries were members. The UK, however, chose not to join. Spain joined in 1989, the UK in 1990 and Portugal in April 1992. Then in September 1992, the UK and Italy indefinitely suspended their membership of the ERM, but Italy rejoined in November 1996 as part of its bid to join the single European currency (see Section 15.6). Austria joined in 1995, Finland in 1996 and Greece in 1998. By the time the ERM was replaced by the single currency in 1999, only Sweden and the UK were outside the ERM.

Features of the ERM

Under the system, each currency was given a central exchange rate with each of the other ERM currencies in a grid. However, fluctuations were allowed from the central rate within specified bands. For most countries these bands were set at ± 2.25 per cent. The central rates could be adjusted from time to time by agreement, thus making the ERM an 'adjustable peg' system. All the currencies floated jointly with currencies outside the ERM.

If a currency approached the upper or lower limit against *any* other ERM currency, intervention would take place to maintain the currencies within the band. This would take the form of central banks in the ERM selling the strong currency and buying the weak one. It could also involve the weak currency countries raising interest rates and the strong currency countries lowering them.

The ERM in practice. In a system of pegged exchange rates, countries should harmonise their policies to avoid excessive currency misalignments and hence the need for large devaluations or revaluations. There should be a convergence of their economies: they should be at a similar point on the business cycle and have similar inflation rates and interest rates.

The ERM in the 1980s. In the early 1980s, however, French and Italian inflation rates were persistently higher than German rates. This meant that there had to be several realignments (devaluations and revaluations). After 1983 realignments became less frequent, and then from 1987 to 1992 they ceased altogether. This was due to a growing convergence of members' internal policies.

By the time the UK joined the ERM in 1990, it was generally seen by its existing members as being a great success. It had created a zone of currency stability in a world of highly unstable exchange rates, and had provided the necessary environment for the establishment of a truly common market by the end of 1992.

Definitions

Adjustable peg A system whereby exchange rates are fixed for a period of time, but may be devalued (or revalued) if a deficit (or surplus) becomes substantial.

Managed floating A system of flexible exchange rates, but where the government intervenes to prevent excessive fluctuations or even to achieve an unofficial target exchange rate.

Exchange rate mechanism (**ERM**) A semi-fixed system whereby participating EU countries allow fluctuations against each other's currencies only within agreed bands. Collectively they float freely against all other currencies.

Crisis in the ERM. Shortly after the UK joined the ERM, however, strains began to show. The reunification of Germany involved considerable reconstruction in the eastern part of the country. Financing this reconstruction was causing a growing budget deficit. The Bundesbank (the German central bank) thus felt obliged to maintain high interest rates in order to keep inflation in check. At the same time, the UK was experiencing a massive current account deficit (partly the result of entering the ERM at what many commentators argued was too high an exchange rate). It was thus obliged to raise interest rates in order to protect the pound, despite the fact that the economy was sliding rapidly into recession. The French franc and Italian lira were also perceived to be overvalued, and there were the first signs of worries as to whether their exchange rates within the ERM could be retained.

At the same time, the US economy was moving into recession and, as a result, US interest rates were cut. This led to a large outflow of capital from the USA. With high German interest rates, much of this capital flowed to Germany. This pushed up the value of the German mark and with it the other ERM currencies.

In September 1992, things reached crisis point. First the lira was devalued. Then two days later, on 'Black Wednesday' (16 September), the UK and Italy were forced to suspend their membership of the ERM: the pound and the lira were floated. At the same time, the Spanish peseta was devalued by 5 per cent.

Pause for thought

Under what circumstances may a currency bloc like the ERM (a) help to prevent speculation; (b) aggravate the problem of speculation?

Turmoil returned in the summer of 1993. The French economy was moving into recession and there were calls for cuts in French interest rates. But this was only possible if Germany was prepared to cut its rates too, and it was not. Speculators began to sell francs and it became obvious that the existing franc/mark parity could not be maintained. In an attempt to rescue the ERM, the EU finance ministers agreed to adopt very wide ± 15 per cent bands. The result was that the franc and the Danish krone depreciated against the mark.

A return of calm. The old ERM appeared to be at an end. The new ± 15 per cent bands hardly seemed like a 'pegged' system at all. However, the ERM did not die. Within months, the members were again managing to keep fluctuations within a very narrow range (for most of the time, within ± 2.25 per cent!). The scene was being set for the abandonment of separate currencies and the adoption of a single currency: the euro.

The Maastricht Treaty and the road to the single currency

The ERM was conceived as a stage on the road to complete economic and monetary union (EMU) of member states. Details of the path towards EMU were finalised in the Maastricht Treaty, which was signed in February 1992. The timetable for EMU involved the adoption of a single currency by 1999 at the latest.

Before they could join the single currency, member states were obliged to achieve convergence of their economies. Each country had to meet five convergence criteria.

- Inflation: should be no more than 1.5 per cent above the average inflation rate of the three countries in the EU with the lowest inflation.
- Interest rates: the rate on long-term government bonds should be no more than 2 per cent above the average of the three countries with the lowest inflation.
- Budget deficit: should be no more than 3 per cent of GDP.
- National debt: should be no more than 60 per cent of GDP.
- Exchange rates: the currency should have been within the normal ERM bands for at least two years with no realignments or excessive intervention.

Before the launch of the single currency, the Council of Ministers had to decide which countries had met the convergence criteria and would thus be eligible to form a **currency union** by fixing their currencies permanently to the euro. Their national currencies would effectively disappear.

At the same time a European System of Central Banks (ESCB) would be created, consisting of a European Central Bank (ECB) and the central banks of the participating member states. The ECB would be independent, both from governments and from EU political institutions. It would operate the monetary policy on behalf of the countries that had adopted the single currency.

Definition

Currency union A group of countries (or regions) using a common currency.

Recap

1. One means of achieving greater currency stability is for a group of countries to peg their internal exchange rates and yet float jointly with the rest of the world. The exchange rate mechanism of the EU (ERM) was an example. Members' currencies were allowed to fluctuate against other member currencies within a band. The band was ±2.25 per cent for the majority of the ERM countries until 1993.
2. The need for realignments seemed to have diminished in the late 1980s as greater convergence was achieved between the members' economies. However, growing strains in the system in the early 1990s led to a crisis in September 1992. The UK and Italy left the ERM. There was a further crisis in July 1993 and the bands were widened to ±15 per cent.
3. Thereafter, as convergence of the economies of ERM members increased, fluctuations decreased and remained largely within ±2.25 per cent.
4. The ERM was seen as an important first stage on the road to complete economic and monetary union (EMU) in the EU.
5. The Maastricht Treaty set out a timetable for achieving EMU. This would culminate in the creation of a currency union: a single European currency with a common monetary policy operated by an independent European Central Bank.

15.6 ECONOMIC AND MONETARY UNION (EMU) IN EUROPE

Do countries benefit from using the euro?

Birth of the euro

In March 1998, the European Commission ruled that 11 of the 15 member states were eligible to proceed to EMU in January 1999. The UK and Denmark were to exercise an opt-out negotiated at Maastricht, and Sweden and Greece failed to meet one or more of the convergence criteria. (Greece joined the euro in 2001.)

The euro came into being on 1 January 1999, but euro banknotes and coins were not introduced until 1 January 2002. In the meantime, national currencies continued to exist alongside the euro, but at irrevocably fixed rates. The old notes and coins were withdrawn a few weeks after the introduction of euro notes and coins.

In May 2004, ten new members joined the EU, in January 2007 another two and in July 2013 another one. Under the Maastricht Treaty, they should all make preparations for joining the euro by meeting the convergence criteria and being in a new version of the exchange rate mechanism with a wide exchange rate band. Under ERM II, euro candidate countries must keep their exchange rates within ±15 per cent of a central rate against the euro. Estonia, Lithuania and Slovenia were the first to join ERM II in June 2004 with Latvia, Cyprus, Malta and Slovakia following in 2005. Slovenia adopted the euro in 2007, Malta and Cyprus in 2008, Slovakia in 2009, Estonia in 2011, Latvia in 2014, Lithuania in 2015 and Croatia in 2023, making a total of 20 countries using the euro.

Advantages of the single currency

EMU has several major advantages for its members.

Elimination of the costs of converting currencies. With separate currencies in each of the EU countries, costs were incurred each time one currency was exchanged into another. The elimination of these costs, however, was probably the least important benefit from the single currency. The European Commission estimated that the effect was to increase the GDP of the countries concerned by an average of only 0.4 per cent and perhaps less so for countries with well-developed financial markets.

Increased competition and efficiency. Despite the advent of the single market, large price differences remained between member states. Not only has the single currency eliminated the need to convert one currency into another (a barrier to competition), but it has brought more transparency in pricing, and has put greater downward pressure on prices in high-cost firms and countries.

Elimination of exchange-rate uncertainty (between the members). Removal of exchange rate uncertainty has helped to encourage trade between the eurozone countries. Perhaps more importantly, it has encouraged investment by firms that trade between these countries, given the greater certainty in calculating costs and revenues from such trade.

KI 15 p 80

Increased inward investment. Investment from the rest of the world is attracted to a eurozone of around 350 million inhabitants, where there is no fear of internal currency movements. As well as potentially increasing its global share of FDI, by removing the uncertainty arising from possible exchange-rate volatility between its 20 members, the eurozone countries may

experience less volatility in FDI flows than would otherwise be the case.

Lower inflation and interest rates. A single monetary policy forces convergence in inflation rates (just as inflation rates are very similar between the different regions within a country). With the ECB being independent from short-term political manipulation, this has resulted in a low average inflation rate in the eurozone countries. This, in turn, has helped to convince markets that the euro will be strong relative to other currencies. The result is lower long-term rates of interest. This, in turn, further encourages investment in the eurozone countries, both by member states and by the rest of the world.

Opposition to EMU

European monetary union has, however, attracted considerable criticism. Many 'Eurosceptics' see within it a surrender of national political and economic sovereignty. Their arguments are essentially ones of principle.

Others, including those more sympathetic to monetary union in principle, raise concerns about the design of the monetary and fiscal systems within which monetary union operates – a design that, in principle, can be amended (see Boxes 13.3 and 13.5).

We begin with those arguments against EMU in principle.

Arguments against EMU in principle

The lack of national currencies. This can be a serious problem if an economy is at all out of harmony with the rest of the eurozone. For example, if countries such as Greece and Spain have lower productivity or higher rates of inflation (due, say, to greater cost-push pressures), then how are they to make their goods competitive with the rest of the Union? With separate currencies these countries could allow their currencies to depreciate. With a single currency, however, they could become depressed 'regions' of Europe, with rising unemployment and all the other regional problems of depressed regions *within* a country.

TC 15 p 287

Pause for thought

How might multiplier effects (the principle of cumulative causation) lead to prosperous regions becoming more prosperous and less prosperous regions falling even further behind?

Proponents of EMU argue that it is better to tackle the problem of high inflation or low productivity in such countries by the disciplines of competition from other EU countries, than merely to feed that inflation by keeping separate currencies and allowing periodic depreciations, with all the uncertainty that they bring.

What is more, the high-inflation countries tend to be the poorer ones with lower wage levels (albeit faster wage *increases*). With higher mobility of labour and capital as the single market develops, resources are likely to be attracted to such countries. This could help to narrow the gap between the richer and poorer member states.

The critics of EMU counter this by arguing that labour is relatively immobile, given cultural and language barriers. Thus an unemployed worker in Dublin could not easily move to a job in Turin or Helsinki. What the critics are arguing here is that the EU is not an **optimal currency area** (see Box 15.5).

Loss of separate monetary policies. Perhaps the most significant criticism is that the same central bank rate of interest must apply to all eurozone countries: the 'one-size-fits-all' problem. The trouble is that while some countries might require a lower rate of interest in order to ward off recession (such as Portugal, Ireland and Greece in 2010–11), others might require a higher one to prevent inflation.

The greater the divergence between economies within the eurozone, the greater this problem becomes. It was hoped, however, that, with common fiscal rules and free trade, these divergences would diminish over time. Eurosceptics, however, argue that there has been increasing *divergence* between members, especially in terms of the size of government deficits and debt and the confidence of the financial markets in countries' ability to tackle these deficits.

Asymmetric shocks. A third and related problem for members of a single currency occurs in adjusting to a shock when that shock affects members to different degrees. These are known as **asymmetric shocks**. For example, the banking crisis affected the UK more severely than other countries, given that London is a global financial centre. This problem is more serious the less the factor mobility between member countries and the less the price flexibility within member countries are.

Definitions

Optimal currency area The optimal size of a currency area is one that maximises the benefits from having a single currency relative to the costs. If the area were to be increased or decreased in size, the costs would rise relative to the benefits.

Asymmetric shocks Shocks (such as an oil price increase or a recession in another part of the world) that have different-sized effects on different industries, regions or countries.

BOX 15.5 OPTIMAL CURRENCY AREAS EXPLORING ECONOMICS

When it pays to pay in the same currency

Imagine that each town and village used a different currency. Think how inconvenient it would be having to keep exchanging one currency into another, and how difficult it would be working out the relative value of items in different parts of the country.

Clearly there are benefits of using a common currency, not only within a country but across different countries. The benefits include greater transparency in pricing, more open competition, greater certainty for investors and the avoidance of having to pay commission when you change one currency into another. There are also the benefits from having a single monetary policy if that is delivered in a more consistent and effective way than by individual countries.

So why not have a single currency for the whole world? The problem is that the bigger a single-currency area gets, the more likely the conditions are to diverge in the different parts of the area. Some parts may have high unemployment and require expansionary policies. Others may have low unemployment and suffer from inflationary pressures. They may require *contractionary* policies.

What is more, different members of the currency area may experience quite different shocks to their economies, whether from outside the union (e.g. a fall in the price of one of their major exports) or from inside (e.g. a prolonged strike). These 'asymmetric shocks' would imply that different parts of the currency area should adopt different policies. But with a common monetary policy and hence common interest rates, and with no possibility of devaluation/revaluation of the currency of individual members, the scope for separate economic policies is reduced.

The costs of asymmetric shocks (and hence the costs of a single-currency area) will be greater the less the mobility of labour and capital and the less the flexibility of prices and wage rates are, and the fewer alternative policies that can be turned to (such as fiscal and regional policies) there are.

So is the eurozone an optimal currency area? Certainly strong doubts have been raised by many economists.

- Labour is relatively immobile.
- There are structural differences between the member states.
- The transmission effects of interest rate changes are different between the member countries. This can arise, for example, if countries' private sectors are exposed differently to debt.
- Exports to countries outside the eurozone account for different proportions of the members' GDP and thus their economies are affected differently by a change in the rate of exchange of the euro against other currencies.
- Wage rates are relatively inflexible.
- Under the Stability and Growth Pact and the Fiscal Compact (see Box 13.3), the scope for using discretionary fiscal policy is curtailed, except in times of severe economic difficulty (e.g. during the COVID-19 pandemic), when an escape clause can be activated.

This does not necessarily mean, however, that the costs of having a single European currency outweigh the benefits. Also, the problems outlined above should decline over time as the single market develops. Finally, the problem of asymmetric shocks can be exaggerated. European economies are highly diversified; there are often more differences *within* economies than between them. Thus shocks are more likely to affect different industries or localities than whole countries. Changing the exchange rate, if that were still possible, would hardly be an appropriate policy in these circumstances.

Why is a single currency area likely to move towards becoming an optimal currency area over time?

Undertake a literature search on work looking at whether the eurozone is an optimal currency area. Write a short review of the relevant literature you discover.

Even when shocks are uniformly felt in the member states, however, there is still the problem that policies adopted centrally will have different impacts on each country. This is because the transmission mechanisms of economic policy (i.e. the way in which policy changes impact on economic variables like growth and inflation) vary across countries.

Criticisms of the current design of EMU

There are others who are critical of the design of EMU but who argue that, with appropriate changes, the problems could be significantly reduced.

Monetary policy. In the case of monetary policy, it was argued that the ECB's remit made it inflation-averse. Its focus on price stability with an inflation target of below, but close to, 2 per cent over the medium term was said to make it more 'hawkish': i.e. less proactive in response to economic downturns, particularly if accompanied by a persistence in inflation. For example, in the early 2010s, in the aftermath of the financial crisis (see Box 13.5), the ECB was more cautious in relaxing monetary policy than the US Federal Reserve, the Bank of England and the Bank of Japan. The reason was that inflation, although falling after 2011, remained above target until 2013.

However, in 2021, following a review of its monetary policy strategy, the ECB adopted a symmetric medium-term target of 2 per cent. It therefore remains to be seen if this addresses the concerns over the hawkishness of the ECB. Certainly, it was much slower to raise interest rates than the Bank of England or the Fed in 2022 as inflation rose following COVID supply-chain disruptions and the Russian invasion of Ukraine.

Critics also point to the underlying weakness of a single currency operating alongside separate *national* government debt issues. The greater the divergence of the eurozone countries, in terms of growth, inflation, deficits, debt and the proportions of debt securities maturing in the short term is, the greater this problem becomes.

KI 1 p 5

Fiscal policy. Under the Stability and Growth Pact (SGP) (which applies to the whole EU), countries were supposed to keep public-sector deficits below 3 per cent of GDP and their stocks of debt below 60 per cent of GDP (see Box 13.3). However, the Pact was not rigidly enforced. Furthermore, because the rules allowed for discretion in times of recession, deficits and debt rose sharply in the late 2000s (see Table 13.1 on page 376).

Subsequently, efforts have been made to change the framework within which national governments make their fiscal choices. The result is the Fiscal Compact, signed in March 2012 (see Box 13.3). This reaffirmed the SGP's excessive deficit rules, but added other requirements. For example, EU countries would now be required to keep *structural deficits* (i.e. budget deficits that would exist even if economies were operating at their potential output level) at or below 0.5 per cent of GDP and tougher penalties would be imposed on countries breaking the rules.

There are those who argue that, for eurozone members to benefit fully from monetary union, tighter fiscal rules alone are insufficient. Instead, they advocate greater fiscal harmonisation. In other words, the problem, they say, is one of incomplete integration.

Pause for thought

By what means would a depressed country in an economic union with a single currency be able to recover? Would the market provide a satisfactory solution to the problems or would (union) government intervention be necessary and, if so, what form could that intervention take?

Future of the euro

When Lithuania adopted the euro on 1 January 2015 it became the nineteenth country to do so. Yet debates around the advantages and disadvantages of EMU intensified as the ongoing Greek debt crisis raised the prospect of Greece's exit from the euro (Grexit).

The Greek crisis

The perilous state of Greece's public finances had already seen two international bailouts agreed. These involved the IMF, the European Commission and the ECB – the so-called 'Troika' – and were worth €240 billion. However, these loans were contingent on the Greek government undertaking a series of economic measures, including significant fiscal tightening. However, the fiscal austerity measures contributed to a deterioration of the macroeconomic environment.

Matters came to a head at the end of 2014 when the final tranches of the Greek bailout programme were suspended by the Troika. This followed the formation in December 2014 of a Syriza-led Greek government which had fought the election on an anti-austerity platform.

Despite lengthy negotiations between Greece and its international creditors, no agreement on further aid was reached to enable Greece to meet a €1.55 billion repayment to the IMF on 30 June 2015. It thus became the first developed country to default on a loan from the IMF. Meanwhile conditions for Greek citizens continued to deteriorate. In July, the ECB announced that it would not increase its emergency liquidity assurance for the Greek financial system beyond existing levels. Without further credit for an already financially distressed banking system, capital controls were imposed with strict limits on withdrawals from bank accounts.

TC 3 p 10

In August 2015 the Greek government and its international creditors reached an agreement on the terms of a third bailout worth €85 billion over three years. Further austerity measures continued. These were required in order for the periodic release of funds, but they were to place further strains on an already weak economy.

Greece exited its bailout programmes in August 2018 and some progress has been made in easing repayment terms. However, there remain concerns that these terms are still damaging the Greek economy. Some, including the IMF, argue that although Greece was permitted in 2019 to re-enter debt markets, Greece's public finances may still not be sustainable in the long term, particularly given the additional pressures arising from the COVID-19 pandemic. If so, they argue, only debt relief would ensure the long-term sustainability of Greece's public finances.

For the time being at least, Grexit had been avoided. However, the Greek debt crisis raises important questions about the conditions under which it would be beneficial for other EU member states to join the euro or for existing members to exit.

The single currency and gains from trade

The benefits from a country being a member of a single currency are greater the more it leads to trade with other members of the single currency. Table 15.2 shows for a sample of eurozone countries the proportion of their exports and imports to and from other EU member states. From the table we can see that about 60 per cent of trade in the EU is between member states. However, there are considerable differences in the importance of intra-EU trade for member states.

Table 15.2 Intra-European Union trade (average 2002–23)

	Intra-EU trade (% of total trade)			Value of intra-EU trade relative to extra-EU trade
	Exports	Imports	Exports and imports	
EU-27	60.1	60.3	60.2	1.51
Eurozone (19 countries)	58.7	58.8	58.7	1.42
Luxembourg	79.4	78.5	79.1	3.77
Estonia	69.2	74.3	71.9	2.56
Austria	68.9	77.0	73.0	2.71
Portugal	69.5	72.9	71.4	2.50
Belgium	65.9	62.4	64.2	1.79
Finland	53.0	65.0	59.0	1.44
France	54.3	63.8	59.4	1.46
Spain	61.7	56.5	58.8	1.43
Germany	53.9	60.7	57.0	1.32
Netherlands	68.7	42.7	56.4	1.29
Cyprus	44.5	59.9	56.1	1.28
Italy	52.6	54.9	53.7	1.16
Greece	53.2	53.2	53.1	1.13
Ireland	41.4	33.2	38.3	0.62

Note: Data from 2022 based on forecasts
Source: Based on data from *AMECO Database* (European Commission, DGECFIN)

On the basis of intra-EU trade, countries like Ireland, Greece and Italy might appear to have least to gain from being part of a single currency with other EU nations. But, we need to consider other factors too. The theory of optimal currency areas (see Box 15.5) suggests, for example, that the degree of convergence between economies and the flexibility of labour markets are important considerations for countries considering the costs of relinquishing their national currency.

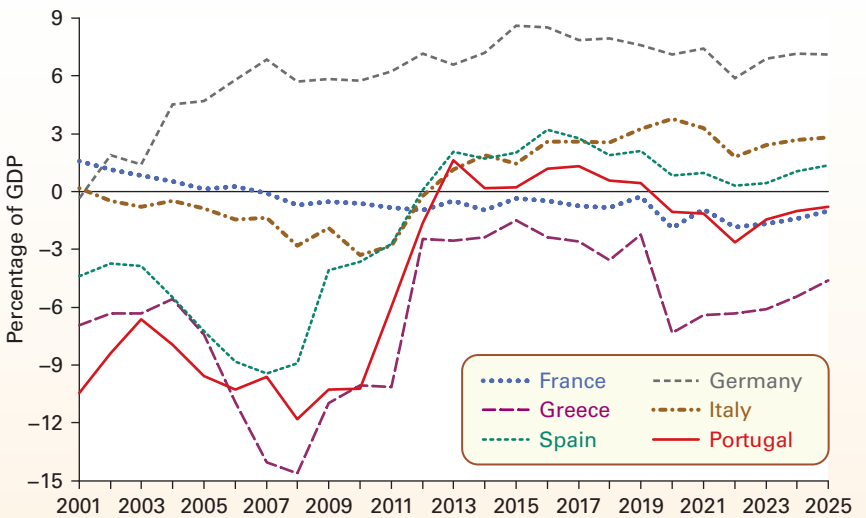
Convergence or divergence?

The more similar economies are, the more likely it is that they will face similar or symmetric shocks which can be accommodated by a common monetary policy. Furthermore, greater wage flexibility and mobility of labour provide mechanisms for countries within a single currency to remain internationally competitive.

However, there remain considerable differences in the macroeconomic performance of eurozone countries, reflecting continuing differences in the structures of their economies. Some of these were exacerbated by the financial crisis of the late 2000s and the subsequent deterioration of the macroeconomic environment.

Among these differences is the contrasting trade positions of eurozone economies. This is illustrated in Figure 15.6 which shows the current account balances of selected eurozone economies since 2001. In the ten-year period from 2001 to 2010, Portugal, Greece and Spain ran large current account deficits, averaging 10, 9 and 6 per cent of GDP respectively. By contrast,

Figure 15.6 Current account balance of selected eurozone economies



Note: Figures from 2022 based on forecasts
Source: Based on data in *AMECO Database* (European Commission, DGECFIN)

Germany ran a current account surplus of around 4 per cent of its GDP. In more recent years, however, the divergences in current account balances have narrowed.

In the absence of nominal exchange-rate adjustments, countries like Greece and Spain, looking to a fall in the *real* exchange rate to boost competitiveness, need to have relatively lower rates of price inflation (see page 459 on nominal and real exchange rates). Therefore, in a single currency, productivity growth and wage inflation take on even greater importance in determining a country's competitiveness. The competitive position of countries will deteriorate if wage growth *exceeds* productivity growth. If this happens, unit labour costs (labour costs per unit of output) will increase.

Figure 15.7 shows how in the period from 2001 to 2008 labour costs increased at an average of between 3 and 4 per cent per annum in Greece, Spain and Italy compared with close to zero in Germany. This, other things being equal, put these countries at a competitive disadvantage.

The euro and the fiscal framework

The discussion so far highlights the importance of economic convergence in affecting the benefits and costs of being a member of the euro. Fiscal policy can provide some buffer against asymmetric shocks by enabling transfers of income to those areas experiencing lower rates of economic growth. Therefore, the fiscal framework within which the euro operates is important when considering the future of the euro.

To date, the eurozone has resisted a centralisation of national budgets. In a more centralised (or federal) system we would see automatic income transfers between different regions and countries. A country, say Greece, affected by a negative economic shock would pay less tax revenues and receive more expenditures from a central eurozone budget, whereas in a country, say Germany, experiencing a positive shock the opposite would be the case.

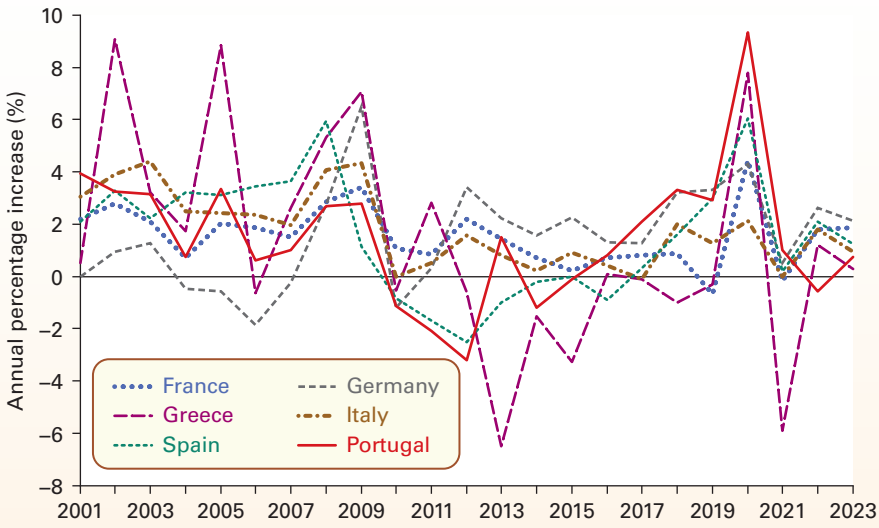
Since national budgets in the eurozone remain largely decentralised, fiscal transfers principally take place *within* countries rather than between them. But this severely limits the use of fiscal policy to offset the effects of negative economic shocks in countries which already have large public-sector deficits and high debt-to-GDP ratios.

To provide maximum flexibility to use fiscal policy, it is important for countries to reduce the stock of public-sector debt as a percentage of annual GDP in times when the economy is growing. As we saw in Section 13.1 (page 378) and in Box 13.1, it will be easier to reduce this percentage if the economy runs a primary surplus. This is when public-sector receipts are greater than public-sector expenditures *excluding* interest payments: the bigger the surplus, the quicker the debt-to-GDP ratio can be reduced. Also, the faster the rate of economic growth and the lower the real rate of interest are, the quicker the ratio can be reduced.[2]

KI 32 p 239

[2]As a rule-of-thumb, the primary surplus-to-GDP ratio required to maintain a given public-sector debt to GDP ratio can be calculated by multiplying the existing debt-to-GDP ratio by the sum of the real rate of interest minus the rate of economic growth (see Box 13.1).

Figure 15.7 Growth in unit labour costs of selected eurozone economies



Notes: Unit labour costs are the ratio of compensation per employee to real GDP per person employed; Figures from 2022 based on forecasts

Source: Based on data in *AMECO Database* (European Commission, DGECFIN May 2022)

Table 15.3 Fiscal indicators in selected eurozone economies

	General government debt-to-GDP (%)		2010–21 averages		
	2010	2021	Primary surplus-to-GDP (%)	Real short-term interest rates (%)	Economic growth (% p.a.)
EU-27	80.5	89.7	−0.8	0.4	1.3
Eurozone	85.9	97.4	−0.9	0.4	1.1
Netherlands	59.2	52.1	−0.7	−0.3	1.3
Ireland	86.2	56.0	−3.6	1.5	6.9
Germany	82.0	69.3	1.0	−0.9	1.5
Austria	82.7	82.8	−0.4	−0.5	1.1
Belgium	100.3	108.2	−0.8	−0.4	1.4
France	85.3	112.9	−2.7	0.3	1.1
Spain	60.5	118.4	−4.0	1.9	0.4
Portugal	100.2	127.4	−0.8	2.8	0.4
Italy	119.2	150.8	0.4	1.8	0.0
Greece	147.5	193.3	−1.6	8.4	−1.8

Source: *AMECO Database* (European Commission, DGECFIN)

Table 15.3 shows the public-sector debt-to-GDP ratios in a sample of eurozone economies (and the UK) in 2010 and 2017 alongside the factors that affect the path of the ratio. The table illustrates considerable differences between countries in the state of their public finances. Therefore, where fiscal policy is left to individual countries, those with an already high debt-to-GDP ratio, such as Greece, Italy and Portugal, will find it considerably more difficult to use fiscal policy to tackle a recession. Consequently, the sustainability of the current decentralised approach to fiscal policy in the eurozone is likely to be crucial in determining the future for the euro and those countries using it.

KI 32 p 239

Recap

1. The euro was born on 1 January 1999. Twelve countries adopted it, having at least nominally met the Maastricht convergence criteria. Euro notes and coins were introduced on 1 January 2002, with the notes and coins of the old currencies withdrawn a few weeks later.
2. The advantages claimed for EMU are that it eliminates the costs of converting currencies and the uncertainties associated with possible changes in inter-EU exchange rates. This encourages more investment, both inward and by domestic firms. What is more, a common central bank, independent from domestic governments, provides the stable monetary environment necessary for a convergence of the EU economies and the encouragement of investment and inter-Union trade.
3. Critics claim, however, that it might make adjustment to domestic economic problems more difficult. The loss of independence in policy making is seen by such people to be a major issue, not only because of the loss of political sovereignty, but also because domestic economic concerns may be at variance with those of the Union as a whole. A single monetary policy is claimed to be inappropriate for dealing with asymmetric shocks. What is more, countries and regions at the periphery of the Union may become depressed unless there is an effective regional policy.
4. The Greek sovereign debt crisis raised concerns about the future of the euro. Considerable differences remain in key macroeconomic indicators. These include differences in the growth of labour productivity and unit labour costs, which are especially significant in the absence of nominal exchange-rate adjustments.
5. There are also considerable differences in the financial health of the public finances of eurozone governments. This is significant because it affects the ability of national governments to use fiscal policy to absorb the adverse economic effects of negative shocks.

15.7 DEBT AND DEVELOPING COUNTRIES

Can their debt burden be lifted?

Perhaps the most serious of all balance of payments problems in the world today is that faced by some of the poorest developing countries. Many of them experience massive financial outflows year after year as a result of having to 'service' debt (i.e. pay interest and make the necessary repayments). Much of this debt has been incurred in their attempts to finance development. Figure 15.8 shows the growth of external debt (as a proportion of national income) that began in the early 1970s.

The oil shocks of the 1970s

In 1973–4 oil prices quadrupled and the world went into recession. Oil imports cost more and export demand was sluggish. The current account deficit of oil-importing developing countries rose from 1.1 percent of GDP in 1973 to 4.3 per cent in 1975.

It was not difficult to finance these deficits, however. The oil surpluses deposited in commercial banks in the industrialised world provided an important additional source of finance. The banks, flush with money and faced with slack demand in the industrialised world, were very willing to lend to developing countries to help them finance continued expansion.

The world recession was short-lived, and with a recovery in the demand for their exports and with their debts being eroded by high world inflation, developing countries found it relatively easy to service these increased debts.

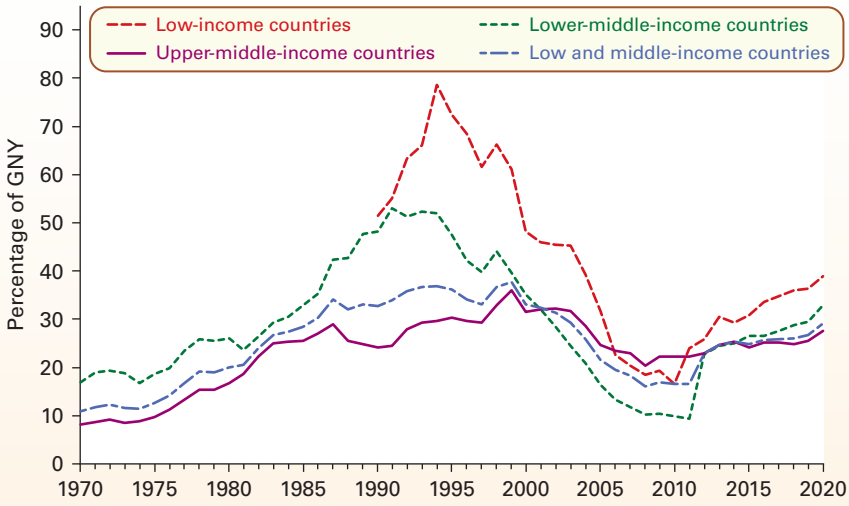
In 1979/80 world oil prices rose again (from $15 to $38 per barrel). This second oil shock, like the first one, caused a large increase in the import bills of developing countries. But the full effects on their economies this time were very much worse, given the debts that had been accumulated in the 1970s and given the policies adopted by the industrialised world after 1979. But why were things so much worse this time?

KI 23 p 173

- The world recession was deeper and lasted longer (1980–3) and, when recovery came, it came very slowly. Developing countries' current account balance of payments deteriorated sharply. This was due both to a marked slowing down in the growth of their exports and to a fall in their export prices.
- The tight monetary policies pursued by the industrialised countries led to a sharp increase in interest rates. This greatly increased developing countries' costs of servicing their debts as can be seen in Figure 15.9 It also led to a sharp fall in inflation, which meant that the debts were not being eroded so rapidly.
- The problem was made worse by the growing proportion of debt that was at variable interest rates.

KI 32 p 239

Figure 15.8 External debt as a percentage of gross national income (GNY)



Source: Based on data from *Databank* (World Bank), series *DT.DOD.DECT.GN.ZS*

This was largely due to the increasing proportion of debt that was in the form of loans from commercial banks.

Figure 15.9 shows how after 1979 debt servicing costs as a proportion of national income rose across developing countries, making it increasingly difficult for them to service their debts. Then in 1982 Mexico, followed by several other countries such as Brazil, Bolivia, Zaire and Sudan, declared that it would have to suspend payments. There was now a debt crisis, which threatened not only the debtor countries, but also the world banking system.

KI 23 p 173

There have been two dimensions to tackling debt problems of developing countries. The first is coping with difficulties in servicing their debt. This usually involves some form of rescheduling of the repayments. The second dimension is dealing with the underlying causes of the problem. Here we will focus on rescheduling.

Rescheduling official loans

Official loans are renegotiated through the 'Paris Club'. Industrialised countries are members of the club, which arranges terms for the rescheduling of their loans to developing countries. Agreements normally involve delaying the date for repayment of loans currently maturing, or spreading the repayments over a longer period of time. Paris Club agreements are often made in consultation with the International Monetary Fund (IMF), which works out a programme for the debtor country to tackle its underlying economic problems.

Several attempts have been made since the mid-1980s to make rescheduling terms more generous, with longer periods before repayments start, longer to repay when they do start, and lower interest rates. In return, the developing countries have had to undertake various 'structural adjustment programmes' supervised by the IMF (see below).

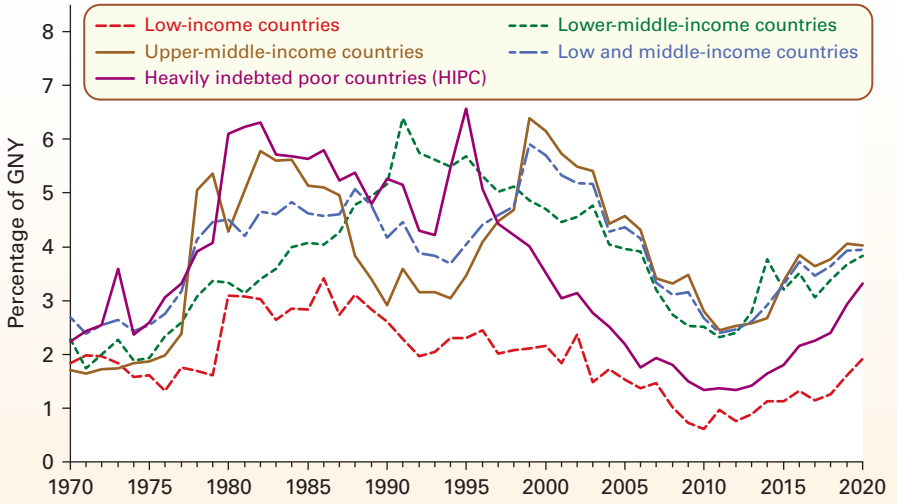
But despite apparent advances made by the Paris Club in making its terms more generous, the majority of low-income countries failed to meet the required IMF conditions, and thus failed to have their debts reduced. What is more, individual Paris Club members were often reluctant to reduce debts unless they were first convinced that other members were 'paying their share'. Nevertheless, some creditor countries unilaterally introduced more generous terms and even cancelled some debts.

The net effect of rescheduling, but only very modest debt forgiveness, can be seen in Figures 15.8 and Figure 15.9. By the mid 1990s average debt service ratios had fallen from the levels of the mid 1980s and yet the ratio of total debt to GNY was higher. There were thus growing calls for the cancellation of debts (see below).

Rescheduling commercial bank loans

After the declarations by Mexico and other countries of their inability to service their debts, there was fear of an imminent collapse of the world banking system. Banks realised that disaster could only be averted by collective action of the banks to reschedule debts. Banks were prepared to reschedule some of the debts and to provide some additional loans in return for debtor countries undertaking structural adjustment (as described below). Additional loans, however, fell

Figure 15.9 Debt servicing costs as a percentage of gross national income (GNY)



Source: Based on data from *Databank* (World Bank), series DT.TDS.DECT.GN.ZS

well short of the amount that was needed. Nevertheless, banks were increasingly setting aside funds to cover bad debt, and thus the crisis for the banks began to recede.

As banks felt less exposed to default, so they became less worried about it and less concerned to negotiate deals with debtor countries. Many of the more severely indebted countries, however, found their position still deteriorating rapidly. What is more, many of them were finding that the IMF adjustment programmes were too painful (often involving deep cuts in government expenditure) and were therefore abandoning them. Thus in 1989 US Treasury Secretary Nicholas Brady proposed measures to *reduce* debt.

Pause for thought

What are the relative advantages and disadvantages to a developing country of rescheduling its debts compared with simply defaulting on them (either temporarily or permanently)?

The *Brady Plan* involved the IMF and the World Bank lending funds to debtor countries to enable them to repay debts to banks. In return for this instant source of liquidity, the banks would have to be prepared to accept repayment of less than the full sum (i.e. they would sell the debt back to the country at a discount). To benefit from such deals, the debtor countries would have to agree to structural adjustment programmes. Several such agreements were negotiated; much of the debt reduction has involved 'debt swaps' of one sort or another (see Case Study 15.9 on the student website).

Dealing with the debt

Structural reforms

The IMF has typically demanded that debtor countries pursue severe structural adjustment programmes before it has been prepared to sanction the rescheduling of debts. Such programmes have included:

- Tight fiscal and monetary policies to reduce government deficits, reduce interest rates and reduce inflation.
- Supply-side reforms to encourage greater use of the market mechanism and greater incentives for investment.
- A more open trade policy and devaluation of the currency in order to encourage more exports and more competition.

These policies, however, often brought extreme hardship as countries were forced to deflate. Unemployment and poverty increased and growth slowed or even became negative. Even though in the long run some developing countries emerged as more efficient and better able to compete in international trade, in the short run the suffering for many was too great to bear. Popular unrest and resentment against the IMF and the country's government led to riots in many countries and a breakdown of law and order.

A more 'complete' structural adjustment would extend beyond simple market liberalisation and tough monetary policies to much more open access to the markets of the rich countries (the subject of much of the Doha Round negotiations: see Box 14.4 on page 430), to more aid and debt relief being channelled into health and education, and to greater research and development in areas that would benefit the poor (e.g. into efficient labour-intensive technology and into new strains of crops that are suitable for countries' specific climate and soil conditions and which do not require large amounts of chemicals).

Debt forgiveness

By the end of the 1990s, the debt burden of many of the poorest countries had become intolerable. Despite portions of their debt being written off under Paris Club terms, the debts of many countries were still rising. Between 1980 and 2000, the debt of sub-Saharan Africa had increased some $3\frac{1}{2}$ times, from $61 billion to $212 billion. Some countries, such as Ethiopia and Mozambique, were spending nearly half their export earnings on merely servicing their debt.

Even with substantial debt rescheduling and some debt cancellation, highly indebted countries were being forced to make savage cuts in government expenditure, much of it on health, education and transport. The consequence was a growth in poverty, hunger, disease and illiteracy. African countries on average were paying four times more to rich countries in debt servicing than they were spending on health and education: it was like a patient giving a blood transfusion to a doctor! The majority of these countries had no chance of 'growing their way out of debt'. The only solution for them was for a more substantial proportion of their debt to be written off.

The heavily indebted poor countries (HIPC) initiative. In 1996 the World Bank and the IMF launched the HIPC initiative. A total of 42 countries, mainly in Africa, were identified as being in need of substantial debt relief. This number was subsequently reduced to 39, most of them in Sub-Saharan Africa. The object of the initiative was to reduce the debts of such countries to 'sustainable' levels by cancelling debts above 200–250 per cent

of GDP (this was reduced to 150 percent in 1999 and to a lower level for five countries).

The HIPC process involved countries passing through two stages. In the first stage, eligible countries must demonstrate a track record of 'good performance'. This means that they had to satisfy the IMF, World Bank and Paris Club that they were undertaking adjustment measures, such as cutting government expenditure and liberalising their markets. It also involved the countries preparing a Poverty Reduction Strategy Paper (PRSP) to show how they would use debt relief to tackle poverty, and especially how this would improve health and education. Once the IMF and World Bank were satisfied that the country was making sufficient progress, the 'decision point' was reached and the level of debt relief would be determined. The country would then enter the second stage.

During the second stage, some interim debt relief is provided. Meanwhile the country must establish a 'sound track record' by implementing policies established at the decision point and based on the PRSP. The length of this stage depends on how long it takes the country to implement the policies. At the end of the second stage, the country reaches the 'completion point' and debts are cancelled by the various creditors, on a pro rata basis, to bring the debt to the sustainable threshold.

In 2006, debt relief for the HIPC countries that reached the completion point was extended under the Multilateral Debt Relief Initiative (MDRI). This involves cancelling multilateral debt incurred before 2004.

By 2020, 37 of the HICPs had reached the completion point and were receiving MDRI relief, while two were at the pre-decision point. As of 2021, $76 billion of debt-service relief over time had been provided under the HIPC programme (in 2017 terms) and a further $43 billion under MDRI. As a result, the debt stocks of the 37 post-decision-point HIPCs had been reduced by over 97 per cent, while poverty-reducing expenditure as a share of government revenue had increased from 42 per cent in 1999 to 46 per cent in 2017.

Despite this substantial relief, the programme has been heavily criticised for taking too long and imposing excessively harsh conditions on the HIPC countries. Then there are the non-HIPCs. Many of these countries also suffer debts which divert a large percentage of their income from poverty relief.

The pandemic and debt

The situation deteriorated significantly in the early 2020s following the outbreak of the COVID-19 pandemic. Developing countries were caught in a policy trilemma: balancing the objectives of increased spending needs from the emergency, limiting the substantial increase in public debt and trying to contain rises in taxes. G20 countries, with the support of the IMF and World Bank, suspended debt service payments from 48 countries between May 2020 and December 2021 through the Debt Service Suspension Initiative (DSSI). These were worth $12.9 billion.

Other measures saw the IMF allocate $100 billion in lending through the Rapid Financing Instrument (RFI) and the Rapid Credit Facility (RCF) and nearly $500 million in debt service relief grants through the Catastrophe Containment and Relief Trust (CCRT). The World Bank increased its operations to $160 billion. The IMF also increased total special drawing rights (SDRs) by $350 billion. However, while these funds come with no conditionality, their distribution is in part determined by the size of economies, which favours richer over poorer economies.

Should all debt be cancelled and aid increased?

In recent years there have been growing calls for the cancellation of debts and a significant increase in aid, especially for the poorest developing countries, many ravaged by war, drought or AIDS. The United Nations has for many years called on wealthy countries to give 0.7 per cent of their gross national income (GNY) in aid.

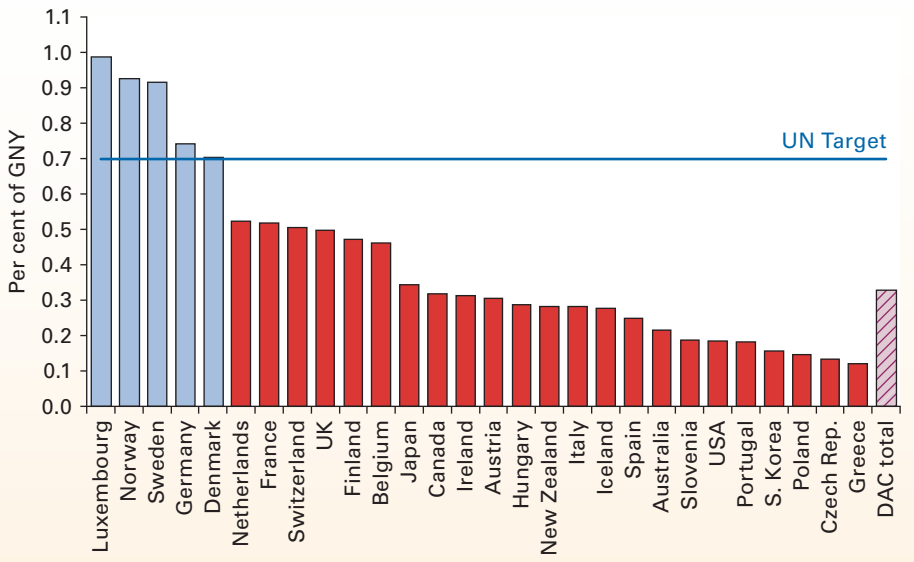
In 2021 the net flow of aid, i.e. official development assistance (ODA), from the 29 donor countries plus European Union institutions which comprise the OECD's Development Assistance Committee (DAC) amounted to $179 billion, including $3.5 billion on providing COVID-19 vaccines. This was a real-terms increase of 4.4 per cent on 2020, though if vaccines are excluded, the increase is a more modest 0.6 per cent. Yet, despite the real-terms increase, ODA in 2021 amounted to a mere 0.33 per cent of the gross national income (GNY) of donor countries – less than half the UN target.

As Figure 15.10 shows, many countries, including the UK, failed to meet the UN target in 2021. The UK had met the target from 2013 to 2019 and in 2015 parliament passed legislation requiring government to meet the 0.7 per cent foreign aid target. However, in 2020 the government announced that, in response to the financial pressures from the COVID-19 pandemic, it would be cutting the aid target to 0.5 per cent. The government stated that it hoped to restore the target when the UK's finances allowed.

The argument against debt cancellation and a substantial increase in aid is that this could represent a moral hazard (see page 83). Once the burden of debt had been lifted and aid had been increased, countries might be tempted to squander the money. It might also encourage them to seek further loans, which might again be squandered.

For this reason, much of the disbursement of funds by donor countries, the IMF and other agencies has been conditional on countries pursuing policies of fiscal restraint and supply-side reform. But also, in recent years, conditionality has required recipient countries

Figure 15.10 Official Development Assistance by DAC countries, per cent of GNY, 2021



Source: Based on data from *Net ODA, OECD Data* (OECD, 2022)

to pursue key poverty-reducing projects, such as health, education, clean water and other basic infrastructure projects. (Case Study 15.20 on the student website examines some of the issues surrounding aid and provides further detail on aid statistics.)

Pause for thought

Should rich countries cancel all debts owed to them by developing countries?

Recap

1. After the 1973 oil crisis many developing countries borrowed heavily in order to finance their balance of trade deficits and to maintain a programme of investment. After the 1979 oil price rises the debt problem became much more serious. There was a world recession and real interest rates were much higher. Debt increased dramatically, and much of it at variable interest rates.
2. Rescheduling can help developing countries cope with increased debt in the short run and various schemes have been adopted by creditor countries and the banks.
3. If the problem is to be tackled, however, then either debts have to be written off – something that banks have been increasingly forced to do – or the developing countries themselves must take harsh corrective measures. The IMF has traditionally favoured 'structural adjustment' policies of deflation and market-orientated supply-side policies. An alternative is to use debt relief and aid to invest in health, education, roads and other infrastructure.
4. In 1996 the World Bank and the IMF launched the HIPC initiative to help reduce the debts of heavily indebted poor countries to sustainable levels. HIPC relief has been criticised, however, for being made conditional on the debtor countries pursuing excessively tough IMF adjustment programmes, for having an excessively long qualifying period and for delays in its implementation. A better approach might be to target debt relief directly at programmes to help the poor.
5. The COVID-19 pandemic had a severe effect on the economies of many poor countries and caused a surge in their debts. The IMF and other international institutions increased support, although the amount of debt cancellation was small compared to HIPC relief.

QUESTIONS

1. Which of the following items are credits on the UK balance of payments and which are debits?
   a. The expenditure by UK tourists on holidays in Greece.
   b. The payment of dividends by foreign companies to investors resident in the UK.
   c. Foreign residents taking out insurance policies with UK companies.
   d. Drawing on reserves.
   e. Investment by UK companies overseas.

(Continued)

2. The table below shows items in the UK's 2000 balance of payments.
   a. Fill in the missing totals for (a) the balance of trade, (b) the current account balance, (c) the portfolio investment balance; and (d) net errors and omissions.
   b. The UK's GDP in 2000 was estimated at £1 098 500 million. Calculate each item on the balance of payments as a percentage of GDP.
   c. Compare the value of each item in £ millions and as percentages of GDP with those for 2021 in Table 15.1.

	£ millions
Current account:	
Balance on trade in goods	−39 479
Balance on trade in services	21 340
Balance of trade	
Income balance	4046
Net current transfers	−9926
Current account balance	
Capital account:	
Capital account balance	393
Financial account:	
Net direct investment	−77 620
Portfolio investment balance	
Other investment balance	2630
Balance of financial derivatives	1553
Reserve assets	−3915
Financial account net flows	24 161
Net errors and omissions	

3. What is the relationship between the balance of trade and the circular flow model?
4. Explain how the current account of the balance of payments is likely to vary with the course of the business cycle.
5. Is it a 'bad thing' to have a deficit on the direct and portfolio investment part of the financial account?
6. Why may credits on a country's short-term financial account create problems for its economy in the future?
7. List some factors that could cause an increase in the credit items of the balance of payments and a decrease in the debit items. What would be the effect on the exchange rate (assuming that it is freely floating)? What effect would these exchange rate movements have on the balance of payments?
8. What policy measures could the government adopt to prevent the exchange rate movements in question 7?
9. What are the major advantages and disadvantages of fixing the exchange rate with a majority currency such as the US dollar?
10. What adverse effects on the domestic economy may follow from (a) a depreciation of the exchange rate, and (b) an appreciation of the exchange rate?
11. What will be the effects on the domestic economy under free-floating exchange rates if there is a rapid expansion in world economic activity? What will determine the size of these effects?
12. Why would banks not be prepared to offer a forward exchange rate to a firm for, say, five years' time?
13. Under what circumstances would the demand for imports be likely to be inelastic? How would an inelastic demand for imports affect the magnitude of fluctuations in the exchange rate?
14. Why are the price elasticities of demand for imports and exports likely to be lower in the short run than in the long run?
15. Assume that the government pursued an expansionary fiscal policy and that the resulting budget deficit led to higher interest rates. What would happen to (a) the current account, and (b) the financial account of the balance of payments? What would be the likely effect on the exchange rate, given a high degree of international financial mobility?
16. Consider the argument that in the modern world of large-scale short-term international financial movements the ability of individual countries to affect their exchange rate is very limited.
17. Why does high international financial mobility and an absence of exchange controls severely limit a country's ability to choose its interest rate?
18. What practical problems are there in achieving a general harmonisation of economic policies between (a) EU countries; (b) the major industrialised countries?
19. What are the causes of exchange rate volatility? Have these problems become greater or lesser in the past ten years? Explain why.
20. Why did the ERM with narrow bands collapse in 1993? Could this have been avoided?
21. Did the exchange rate difficulties experienced by countries under the ERM strengthen or weaken the arguments for progressing to a single European currency?
22. Under what circumstances may a pegged exchange rate system like the ERM (a) help to prevent speculation; (b) aggravate the problem of speculation?
23. By what means would a depressed country in an economic union with a single currency be able to recover? Would the market provide a satisfactory solution or would (union) government intervention be necessary, and if so, what form would the intervention take?
24. Assume that just some of the members of a common market like the EU adopt full economic and monetary union, including a common currency. What are the advantages and disadvantages to those members joining the full EMU and to those not joining?
25. Is the eurozone likely to be an optimal currency area? Is it more or less likely to be so over time? Explain your answer.
26. It is often argued that international convergence of economic indicators is a desirable objective. Does this mean that countries should seek to achieve the same rate of economic growth, monetary growth, interest rates, budget deficits as a percentage of their GDP, etc.?
27. Why is it difficult to achieve international harmonisation of economic policies?
28. To what extent was the debt crisis of the early 1980s caused by inappropriate policies that had been pursued by the debtor countries?

29. Imagine that you are an ambassador of a developing country at an international conference. What would you try to persuade the rich countries to do in order to help you and other poor countries overcome the debt problem? How would you set about persuading them that it was in their own interests to help you?
30. Has increased globalisation been a good thing for national economies?

ADDITIONAL CASE STUDIES ON THE *ESSENTIALS OF ECONOMICS* STUDENT WEBSITE (www.pearsoned.co.uk/sloman)

15.1 **Balance of trade and the public finances.** An examination of countries' budget and balance of trade balances.

15.2 **Making sense of the financial balances on the balance of payments.** An examination of the three main components of the financial account.

15.3 **The UK's balance of payments deficit.** An examination of the UK's persistent trade and current account deficits.

15.4 **Does PPP hold in the long run?** This considers the relationship between inflation rate differentials and movements in sterling.

15.5 **The Gold Standard.** An historical example of fixed exchange rates.

15.6 **A high exchange rate.** This case looks at whether a high exchange rate is necessarily bad news for exporters.

15.7 **The sterling crisis of early 1985.** When the pound fell almost to $1.00.

15.8 **Currency turmoil in the 1990s.** A crisis in Mexico; a rising yen and German mark; a falling US dollar – why did this all happen?

15.9 **The 1997/8 crisis in Asia.** The role played by the IMF.

15.10 **Argentina in crisis.** An examination of the collapse of the Argentinean economy in 2001/2 and subsequent events.

15.11 **The Tobin tax.** An examination of the possible use of small taxes on foreign exchange transactions. The purpose is to reduce currency fluctuations.

15.12 **The euro, the US dollar and world currency markets.** An analysis of the relationship between the euro and the dollar.

15.13 **UK experience of floating.** An historical account of the volatility in sterling since the early 1970s.

15.14 **Using interest rates to control both aggregate demand and the exchange rate.** A problem of one instrument and two targets.

15.15 **The UK Labour government's convergence criteria for euro membership.** An examination of the five tests set by the UK's incoming Labour government in 1997 that needed to be passed before euro membership would have been put to the electorate in a referendum.

15.16 **Debt and the environment.** How high levels of debt can encourage developing countries to damage their environment in an attempt to increase export earnings.

15.17 **2030 Agenda for Sustainable Development.** An overview of the UN's Agenda for Sustainable Development and its associated development goals and targets.

15.18 **The great escape.** This case examines the problem of capital flight from developing countries to rich countries.

15.19 **Swapping debt.** Schemes to convert a developing country's debt into other forms, such as shares in its industries.

15.20 **Economic aid.** Does aid provide a solution to the debt problem?

WEB APPENDIX

5.1 The effectiveness of fiscal and monetary policies. A comparison of the effectiveness of fiscal and monetary policies under fixed and floating exchange rates.

WEBSITES RELEVANT TO PART D

Numbers and sections refer to websites listed in the Websites Appendix at the end of the text and hotlinked from this book's website at www.pearsoned.co.uk/sloman.

- For news articles relevant to Part D, search for the *Sloman Economics News* site or follow the News Items link from the student website or MyLab Economics.
- For general news on countries' trade, balance of payments and exchange rates, see websites in section A, and particularly A1–5, 7–9, 20–26, 31, 35, 36. For articles on various aspects of economic development, see A27, 28. See also links to newspapers worldwide in A38, 39, 43, 44, and the news search feature in Google at A41. See also links to economics news in A42.

(Continued)

- For international data on imports and exports, see section B in site B1. See also H16 > *Documents, data and resources* > *Statistics*. See also trade data in B31, 35, 42, 43.
- For international data on balance of payments and exchange rates, see B1, B31 and OECD Economic Outlook in B21 (also in section 7 of B1).
- For UK data on balance of payments, see B3: search for *Pink Book* (annual) and *United Kingdom Economic Accounts* (quarterly). See also B34. For EU data, see B38 and 47.
- For exchange rates, see A3; B1, 3; F2, 6, 8.
- For data on debt and development, see B24 (search for *Debt* and *World Development Indicators*) and B31.
- For discussion papers on international trade, the balance of payments and exchange rates, see H4 and 7.
- For trade disputes, see H16.
- For information on various preferential trading arrangements, see H20–23.
- For various pressure groups critical of the effects of free trade and globalisation, see H11, 13, 14.
- For EU sites, see G1, 3, 7–14, 16–23.
- For information on trade, debt and developing countries, see H4, 7, 9, 10, 13, 14, 16–19; I9, 21.
- For student resources relevant to Part D, see sites C1–7, 9, 10, 19.

Websites appendix

All the following websites can be accessed from this book's own website (http://www.pearsoned.co.uk/sloman). When you enter the site, click on **Hot Links.** You will find all the following sites listed. Click on the one you want and the 'hot link' will take you straight to it.

The sections and numbers below refer to the ones used in the websites listed at the end of each chapter. Thus if the list contained the number A21, this would refer to the *Conversation* site.

A General news sources

As the title of this section implies, the websites here can be used for finding material on current news issues or tapping into news archives. Most archives are offered free of charge. However, some do require you to register. As well as key UK and American news sources, you will also notice some slightly different places from where you can get your news, such as *The Moscow Times* and *The Japan Times* Check out site numbers 38. *Refdesk,* 43. *Guardian World News Guide* and 44. *Online newspapers* for links to newspapers across the world. Try searching for an article on a particular topic by using site number 41. *Google News Search*.

1. BBC news
2. The Economist
3. The Financial Times
4. The Guardian
5. The Independent
6. ITN
7. The Observer
8. The Telegraph
9. Aljazeera
10. The New York Times
11. Fortune
12. Time Magazine
13. The Washington Post
14. The Moscow Times (English)
15. Pravda (English)
16. Straits Times (Singapore)
17. New Straits Times (Malaysia)
18. The Scotsman
19. The Herald
20. Euromoney
21. The Conversation
22. Market News International
23. Bloomberg Businessweek
24. International Business Times
25. CNN Money
26. Vox (economic analysis and commentary)
27. Asia News Network
28. allAfrica.com
29. Greek News Sources (English)
30. France 24 (English)
31. Euronews
32. Australian Financial Review
33. Sydney Morning Herald
34. The Japan Times
35. Reuters
36. Bloomberg
37. David Smith's Economics UK.com
38. Refdesk (links to a whole range of news sources)
39. Newspapers and Magazines on World Wide Web
40. Yahoo News Search
41. Google News Search
42. ABYZ news links
43. Guardian World News Guide
44. Online newspapers

B Sources of economic and business data

Using websites to find up-to-date data is of immense value to the economist. The data sources below offer you a range of specialist and non-specialist data information. Universities have free access to the *UK Data Service* site (site 35 in this set), which is a huge database of statistics. Site 3, the *Office for National Statistics* is the UK's main official data site and datasets can be downloaded as Excel files. The Economics Network's *Economic data freely available online* (site 1) gives links to various sections in over 40 UK and international sites.

1. Economics Network gateway to economic data
2. Office for Budget Responsibility
3. Office for National Statistics
4. Data Archive (Essex)
5. Bank of England Statistical Interactive Database
6. UK official statistics (GOV.UK)
7. Nationwide House Prices Site
8. House Web (data on housing market)
9. Economist global house price data
10. Halifax House Price Index
11. House prices indices from ONS

12. Penn World Table
13. Economist economic and financial indicators
14. FT market data
15. Economagic
16. Groningen Growth and Development Centre
17. AEAweb: Resources for economists on the Internet (RFE): data
18. Joseph Rowntree Foundation
19. OECD iLibrary statistics
20. Energy Information Administration
21. OECDStat
22. CIA world statistics site (World Factbook)
23. Millennium Development Goal Indicators Database
24. World Bank Data
25. Federal Reserve Bank of St Louis, US Economic Datasets (FRED)
26. Ministry of Economy, Trade and Industry (Japan)
27. Financial data from Yahoo
28. DataMarket
29. Index Mundi
30. Knoema: Economics
31. World Economic Outlook Database (IMF)
32. Telegraph shares and markets
33. Key Indicators (KI) for Asia and the Pacific Series (Asia Development Bank)
34. Open data from data.gov.uk (Business and Economy)
35. UK Data Service (incorporating ESDS)
36. BBC News, market data
37. NationMaster
38. Economic Forecasts (European Commission)
39. Business and Consumer Surveys (all EU countries)
40. Gapminder
41. Trading Economics
42. WTO International Trade Statistics database
43. UNCTAD trade, investment and development statistics (UNCTADstat)
44. London Metal Exchange
45. Bank for International Settlements, global nominal and real effective exchange rate indices
46. United Nations: Monthly Bulletin of Statistics
47. AMECO database
48. The Conference Board data
49. Institute for Fiscal Studies: tools and resources
50. European Central Bank (ECB): statistics

C Sites for students and teachers of economics

The following websites offer useful ideas and resources to those who are studying or teaching economics. It is worth browsing through some just to see what is on offer. Try out the first four sites, for starters. The *Internet for Economics* (site 8) is a very helpful tutorial for economics students on using the Internet.

1. The Economics Network
2. Teaching Resources for Undergraduate Economics (TRUE)
3. EconEdLink
4. Studying Economics
5. Economics and Business Education Association
6. Tutor2U
7. Council for Economic Education
8. Dollars and Sense
9. Econoclass: Resources for economics teachers
10. Teaching resources for economists (RFE)
11. METAL – Mathematics for Economics: enhancing Teaching and Learning
12. Federal Reserve Bank of San Francisco: Economics Education
13. Excel in Economics Teaching (from the Economics Network)
14. Economics Online
15. Dr. T's EconLinks: Teaching Resources
16. Online Opinion (Economics)
17. Free to Choose TV from the Idea Channel
18. History of Economic Thought
19. Resources For Economists on the Internet (RFE)
20. Games Economists Play (non-computerised classroom games)
21. Bank of England education resources
22. Why Study Economics?
23. Economic Classroom Experiments
24. Veconlab: Charles Holt's classroom experiments
25. Experiments, games and role play from the Economics Network
26. MIT Open Courseware in Economics
27. EconPort
28. ThoughtCo. – Economics

D Economic models, simulations and classroom experiments

Economic modelling is an important aspect of economic analysis. There are several sites that offer access to a model or simulation for you to use, e.g. *Virtual Chancellor* (where you can play being Chancellor of the Exchequer). Using such models can be a useful way of finding out how economic theory works within a specific environment. Other sites link to games and experiments, where you can play a particular role, perhaps competing with other students.

1. Virtual Chancellor
2. Virtual Factory
3. Interactive simulation models (Economics Web Institute)
4. Classroom Experiments in Economics (Pedagogy in Action)
5. MobLab
6. Economics Network Handbook, Chapter on Simulations, Games and Role-play
7. Experimental Economics Class Material (David J Cooper)
8. Simulations
9. Experimental economics: Wikipedia
10. Software available on the Economics Network site
11. RFE Software
12. Virtual Worlds
13. Veconlab: Charles Holt's classroom experiments
14. EconPort Experiments
15. Denise Hazlett's Classroom Experiments in Macroeconomics
16. Games Economists Play

17. Finance and Economics Experimental Laboratory at Exeter (FEELE)
18. Classroom Expernomics
19. The Economics Network's Guide to Classroom Experiments and Games
20. Economic Classroom Experiments (Wikiversity)

E UK government and UK organisations' sites

If you want to see what a government department is up to, then look no further than the list below. Government departments' websites are an excellent source of information and data. They are particularly good at offering information on current legislation and policy initiatives.

1. Gateway site (GOV.UK)
2. Department for Levelling Up, Housing and Communities
3. Prime Minister's Office
4. Competition & Markets Authority
5. Department for Education
6. Department for International Development
7. Department for Transport
8. Department of Health and Social Care
9. Department for Work and Pensions
10. Department for Business, Energy & Industrial Strategy
11. Environment Agency
12. Former Department of Energy and Climate Change
13. Low Pay Commission
14. Department for Environment, Food & Rural Affairs (Defra)
15. Office of Communications (Ofcom)
16. Office of Gas and Electricity Markets (Ofgem)
17. Official Documents OnLine
18. Office for Budget Responsibility
19. Office of Rail and Road (ORR)
20. The Takeover Panel
21. Sustainable Development Commission
22. Ofwat
23. National Statistics (ONS)
24. List of ONS releases from UK Data Explorer
25. HM Revenue & Customs
26. UK Intellectual Property Office
27. Parliament website
28. Scottish Government
29. Scottish Environment Protection Agency
30. HM Treasury
31. Equality and Human Rights Commission
32. Trades Union Congress (TUC)
33. Confederation of British Industry (CBI)
34. Adam Smith Institute
35. Chatham House
36. Institute for Fiscal Studies
37. Advertising Standards Authority
38. Businesses and Self-employed
39. Campaign for Better Transport
40. New Economics Foundation
41. Financial Conduct Authority
42. Prudential Regulation Authority

F Sources of monetary and financial data

As the title suggests, here are listed useful websites for finding information on financial matters. You will see that the list comprises mainly central banks, both within Europe and further afield. The links will take you to English language versions of non-English speaking countries' sites.

1. Bank of England
2. Bank of England Monetary and Financial Statistics
3. Banque de France (in English)
4. Bundesbank (German central bank)
5. Central Bank of Ireland
6. European Central Bank
7. Eurostat
8. US Federal Reserve Bank
9. Netherlands Central Bank (in English)
10. Bank of Japan (in English)
11. Reserve Bank of Australia
12. Bank Negara Malaysia (in English)
13. Monetary Authority of Singapore
14. Bank of Canada
15. National Bank of Denmark (in English)
16. Reserve Bank of India
17. Links to central bank websites from the Bank for International Settlements
18. The London Stock Exchange

G European Union and related sources

For information on European issues, the following is a wide range of useful sites. The sites maintained by the European Union are an excellent source of information.

1. Business, Economy, Euro: (EC DG)
2. European Central Bank
3. EU official website
4. Eurostat
5. Employment, Social Affairs and Inclusion: (EC DG)
6. Reports, Studies and Booklets on the EU
7. Internal Market, Industry, Entrepreneurship and SMEs: (EC DG)
8. Competition: (EC DG)
9. Agriculture and Rural Development: (EC DG)
10. Energy: (EC DG)
11. Environment: (EC DG)
12. Regional and Urban Policy: (EC DG)
13. Taxation and Customs Union: (EC DG)
14. Education and Training: (EC DG)
15. European Patent Office
16. European Commission
17. European Parliament
18. European Council
19. Mobility and Transport: (EC DG)

20. Trade (EC DG)
21. Maritime Affairs and Fisheries: (EC DG)
22. International Partnerships: (EC DG)
23. Financial Stability, Financial Services and Capital Markets Union (EC DG)

H International organisations

This section casts its net beyond Europe and lists the Web addresses of the main international organisations in the global economy. You will notice that some sites are run by charities, such as Oxfam, while others represent organisations set up to manage international affairs, such as the International Monetary Fund and the United Nations.

1. UN Food and Agriculture Organization (FAO)
2. United Nations Conference on Trade and Development (UNCTAD)
3. International Labour Organization (ILO)
4. International Monetary Fund (IMF)
5. Organisation for Economic Co-operation and Development (OECD)
6. OPEC
7. World Bank
8. World Health Organization (WHO)
9. United Nations (UN)
10. United Nations Industrial Development Organisation (UNIDO)
11. Friends of the Earth
12. Institute of International Finance
13. Oxfam
14. Christian Aid (reports on development issues)
15. European Bank for Reconstruction and Development (EBRD)
16. World Trade Organization (WTO)
17. United Nations Development Programme
18. UNICEF
19. EURODAD – European Network on Debt and Development
20. NAFTA
21. South American Free Trade Areas
22. ASEAN
23. APEC

I Economics search and link sites

If you are having difficulty finding what you want from the list of sites above, the following sites offer links to other sites and are a very useful resource when you are looking for something a little bit more specialist. Once again, it is worth having a look at what these sites have to offer in order to judge their usefulness.

1. Gateway for UK official sites
2. Alta Plana
3. Data Archive Search
4. Inomics: information on economics courses and jobs
5. Ideas from RePEc: bibliographic database
6. Wikidata
7. Portal sites with links to other sites (Economics Network)
8. Economic Links from California State University San Marcos
9. Global goals 2030 (link to economic development resources)
10. Development Data Hub
11. DMOZ Open Directory: Economics (legacy site)
12. Web links for economists from the Economics Network
13. EconData.Net
14. Yale university: 75 Sources of Economic Data, Statistics, Reports, and Commentary
15. Excite Economics Links
16. Resources for Economists
17. Trade Map (trade statistics)
18. Resources for Economists on the Internet
19. UK University Economics Departments
20. Research in Economics Education
21. Development Gateway
22. Data.gov
23. WikiProject: Economics
24. National Bureau of Economic Research links to data sources

J Internet search engines

The following search engines have been found to be useful.

1. Google
2. Bing
3. Whoosh UK
4. Excite
5. Search.com
6. MSN
7. DuckDuckGo
8. Yahoo
9. Ask
10. Lycos
11. Webcrawler
12. Metacrawler: searches several search engines

Key ideas and glossary

KEY IDEAS

1. **Scarcity** is the excess of human wants over what can actually be produced. Because of scarcity, various choices have to be made between alternatives (page 5).
2. **Opportunity cost.** The cost of something measured in terms of what you give up to get it/do it. The best alternative forgone (page 7). (Threshold Concept 1)
3. **Rational decision-making** involves weighing up the marginal benefit and marginal cost of any activity. If the marginal benefit exceeds the marginal cost, it is rational to do the activity (or to do more of it). If the marginal cost exceeds the marginal benefit, it is rational not to do it (or to do less of it) (page 9). (Threshold Concept 2)
4. **Economies suffer from inherent instability.** As a result, economic growth and other macroeconomic indicators tend to fluctuate (page 10). (Threshold Concept 3)
5. **Modelling in economics** involves specifying how one variable (the 'dependent variable') depends on one or more other variables ('independent variables'). It involves 'holding constant' all other variables that might influence the outcome (the *ceteris paribus* assumption). A model can be expressed in words, as a graph, or mathematically in terms of one or more equations. In this book we use mainly verbal descriptions and graphs (page 13). (Threshold Concept 4)
6. **The distinction between nominal and real figures.** Nominal figures are those using current prices, interest rates, etc. Real figures are figures corrected for inflation (page 17). (Threshold Concept 5)
7. **People respond to incentives.** It is important, therefore, that incentives are appropriate and have the desired effect (page 24). (Threshold Concept 6)
8. **Changes in demand or supply cause markets to adjust.** Whenever such changes occur, the resulting 'disequilibrium' will bring an automatic change in prices, thereby restoring equilibrium (i.e. a balance of demand and supply) (page 24).
9. **Government intervention may be able to rectify various failings of the market.** Government intervention in the market can be used to achieve various economic objectives that may not be best achieved by the market. Governments, however, are not perfect and their actions may bring adverse as well as beneficial consequences (page 26). (Threshold Concept 7)
10. **Equilibrium is the point where conflicting interests are balanced.** Only at this point is the amount that demanders are willing to purchase the same as the amount that suppliers are willing to supply. It is a point that will be automatically reached in a free market through the operation of the price mechanism (page 37). (Threshold Concept 8)
11. **People gain from voluntary interaction.** When people buy from or sell to other people, or when they are employed by or employ other people, both parties will gain from the interaction (page 45). (Threshold Concept 9)
12. **Elasticity.** The responsiveness of one variable (e.g. demand) to a change in another (e.g. price). This concept is fundamental to understanding how markets work. The more elastic variables are, the more responsive the market is to changing circumstances (page 49). (Threshold Concept 10)
13. **People's actions are influenced by their expectations.** People respond not just to what is happening now (such as a change in price), but to what they anticipate will happen in the future (page 61). (Threshold Concept 11)
14. **The principle of diminishing marginal utility.** The more of a product a person consumes over a given period of time, the less will be the additional utility gained from one more unit will be (page 72).
15. **People's actions are influenced by their attitudes towards risk.** Many decisions are taken under conditions of risk or uncertainty. Generally, the lower the probability of (or the more uncertain) the desired outcome of an action, the less likely it is that people will undertake the action (page 80).
16. **Output depends on the amount of resources and how they are used.** Different amounts and combinations of inputs will lead to different amounts of output. If output is to be produced efficiently, then inputs should be combined in the optimum proportions (page 94).
17. **The law of diminishing marginal returns.** When increasing amounts of a variable factor are used with a given amount of a fixed factor, there will come a point when each extra unit of the variable factor will produce less extra output than the previous unit (page 95).
18. **The principal–agent problem.** Where people (principals), as a result of a lack of knowledge, cannot ensure that their best interests are served by their agents. Agents may take

advantage of this situation to the disadvantage of the principals (page 117).

19. **Market power benefits the powerful at the expense of others.** When firms have market power over prices, they can use this to raise prices and profits above the perfectly competitive level. Other things being equal, the firm will gain at the expense of the consumer. Similarly, if consumers or workers have market power, they can use this to their own benefit (page 122).
20. **Economic efficiency** is achieved when each good is produced at the minimum cost and where consumers get maximum benefit from their income (page 127).
21. **People often think and behave strategically.** How you think others will respond to your actions is likely to influence your own behaviour. Firms, for example, when considering a price or product change will often take into account the likely reactions of their rivals (page 140).
22. **The fallacy of composition.** What applies in one case will not necessarily apply when repeated in all cases (page 153).
23. **Stocks and flows.** A stock is a quantity of something at a given point in time. A flow is an increase or decrease in something over a specified period of time. This is an important distinction and a common cause of confusion (page 173).
24. **Equity.** Where income or wealth is distributed in a way that is considered to be fair or just. Note that an equitable distribution is not the same as a totally equal distribution and that different people have different views on what is equitable (page 185).
25. **Allocative efficiency in any activity is achieved where any reallocation would lead to a decline in net benefit.** It is achieved where marginal benefit equals marginal cost. Private efficiency is achieved where marginal private benefit equals marginal private cost ($MB = MC$). Social efficiency is achieved where marginal social benefit equals marginal social cost ($MSB = MSC$) (page 195). (Threshold Concept 12)
26. **Markets generally fail to achieve social efficiency.** There are various types of market failure. Market failures provide one of the major justifications for government intervention in the economy (page 195).
27. **General equilibrium.** A situation where all markets are in equilibrium. This situation can be assessed as to whether or not allocative efficiency is achieved. This will depend on whether or not markets are perfect (page 195). (Threshold Concept 13)
28. **Externalities are spillover costs or benefits.** Where these exist, even an otherwise perfect market will fail to achieve social efficiency (page 196).
29. **The free-rider problem.** People are often unwilling to pay for things if they can make use of things other people have bought. This problem can lead to people not purchasing things which would be to the benefit of themselves and other members of society to have (page 199).
30. **The problem of time lags.** Many economic actions can take a long time to take effect. This can cause problems of instability and an inability of the economy to achieve social efficiency (page 203).
31. **Societies face trade-offs between economic objectives.** For example, the goal of faster growth may conflict with that of greater equality; the goal of lower unemployment may conflict with that of lower inflation (at least in the short run). This is an example of opportunity cost: the cost of achieving more of one objective may be achieving less of another. The existence of trade-offs means that policy makers must make choices (page 206).
32. **Balance sheets affect people's behaviour.** The size and structure of governments', institutions' and individuals' liabilities (and assets too) affect economic well-being and can have significant effects on behaviour and economic activity (page 239).
33. **The principle of cumulative causation.** An initial event can cause an ultimate effect that is much larger (page 249). (Threshold Concept 14)
34. **Long-term growth in a country's output depends on growth in the quantity and/or productivity of its resources.** Potential economic growth depends on the country's resources, technology and productivity. This is crucial to understanding what underlies the wealth of nations and why some countries have faster growth rates than others (page 287). (Threshold Concept 15)
35. **Goodhart's Law.** Controlling a symptom (i.e. an indicator) of a problem will not cure the problem. Instead, the indicator will merely cease to be a good indicator of the problem (page 400).
36. **The law of comparative advantage.** Provided opportunity costs of various goods differ in two countries, both of them can gain from mutual trade if they specialise in producing (and exporting) those goods that have relatively low opportunity costs compared with the other country's (page 422).

GLOSSARY

Absolute advantage A country has an absolute advantage over another in the production of a good if it can produce it with less resources than the other country can.

Accelerationist hypothesis The theory that unemployment can only be reduced below the natural level at the cost of accelerating inflation.

Accelerator coefficient The level of induced investment as a proportion of a rise in national income: $\alpha = I_i/\Delta Y$.

Accelerator theory The *level* of investment depends on the *rate of change* of national income, and as a result tends to be subject to substantial fluctuations.

Active balances Money held for transactions and precautionary purposes.

Actual growth The percentage annual increase in national output actually produced.

Ad valorem tariffs Tariffs levied as a percentage of the price of the import.

Ad valorem tax A tax on a good levied as a percentage of its value. It can be a single-stage tax or a multi-stage tax (such as VAT).

Adaptive expectations Where people adjust their expectations of inflation in the light of what has happened to inflation in the past.

Adjustable peg A system whereby exchange rates are fixed for a period of time, but may be devalued (or revalued) if a deficit (or surplus) becomes substantial.

Adverse selection A market process whereby either buyers, sellers or products with certain unobservable characteristics (e.g. high risk or low quality) are more likely to enter the market at the current market price. In the context of insurance, it is the tendency of those at greatest risk to take out insurance.

Aggregate demand Total spending on goods and services made in the economy. It consists of four elements, consumer spending (C), investment (I), government spending (G) and the expenditure on exports (X), less any expenditure on imports of goods and services (M): $AD = C + I + G + X - M$.

Aggregate demand for labour curve A curve showing the total demand for labour in the economy at different levels of real wage rates.

Aggregate supply The total amount that firms plan to supply at any given level of prices.

Aggregate supply of labour curve A curve showing the total number of people willing and able to work at different average real wage rates.

Allocative efficiency A situation where the current combination of goods produced and sold gives the maximum satisfaction for each consumer at their current levels of income. Note that a redistribution of income would lead to a different combination of goods that was allocatively efficient.

Allocative role for government intervention Interventions by government to affect the allocation of resources in consumption and/or production.

Appreciation A rise in the free-market exchange rate of the domestic currency with foreign currencies.

Arc elasticity The measurement of elasticity between two points on a curve.

Assets Possessions of an individual or institution, or claims held on others.

Asymmetric information Where one party in an economic relationship (e.g. an agent) has more information than another (e.g. the principal).

Asymmetric shocks Shocks (such as an oil price increase or a recession in another part of the world) that have different-sized effects on different industries, regions or countries.

Automatic fiscal stabilisers Tax revenues that rise and government expenditure that falls as national income rises. The more they change with income, the bigger the stabilising effect on national income.

Average cost pricing or **mark-up pricing** Where firms set the price by adding a profit mark-up to average cost.

Average fixed cost Total fixed cost per unit of output: $AFC = TFC/Q$.

Average (or 'mid-point') formula for price elasticity of demand ΔQ_D/average $Q_D \div \Delta P$/average P.

Average physical product Total output (TPP) per unit of the variable factor in question: $APP = TPP/Q_v$.

Average propensity to consume The proportion of income that is spent (C/Y).

Average rate of income tax Income taxes as a proportion of a person's total (gross) income: T/Y.

Average revenue Total revenue per unit of output. When all output is sold at the same price, average revenue will be the same as price: $AR = TR/Q = P$.

Average (total) cost Total cost (fixed plus variable) per unit of output: $AC = TC/Q = AFC + AVC$.

Average variable cost Total variable cost per unit of output: $AVC = TVC/Q$.

Backwards induction A process by which firms think through the most likely outcome in the last period of competition and then work backwards step by step thinking through the most likely outcomes in earlier periods of competition.

Balance of payments account A record of the country's transactions with the rest of the world. It shows the country's payments to or deposits in other countries (debits) and its receipts or deposits from other countries (credits). It also shows the balance between these debits and credits under various headings.

Balance of payments on current account The balance on trade in goods and services plus net investment income and current transfers.

Balance of trade Exports of goods and services minus imports of goods and services. If exports exceed imports, there is a 'balance of trade surplus' (a positive figure). If imports exceed exports, there is a 'balance of trade deficit' (a negative figure).

Balance on trade in goods Exports of goods minus imports of goods.

Balance on trade in goods and services (or balance of trade) Exports of goods and services minus imports of goods and services.

Balance on trade in services Exports of services minus imports of services.

Balance sheet A record of the stock of assets and liabilities of an individual or institution.

Balance sheet effects The effects on spending behaviour, such as consumer spending, that arise from changes in the composition or value of net worth.

Balance sheet recession An economic slowdown or recession caused by private-sector agents looking to improve their financial well-being by increasing their saving and/or paying down debt.

Balancing item (in the balance of payments) A statistical adjustment to ensure that the two sides of the balance of payments account balance. It is necessary because of errors in compiling the statistics.

Bank bills Bills that have been accepted by another institution and hence insured against default.

Bank (or deposits) multiplier The number of times greater the expansion of bank deposits is than the additional liquidity in banks that causes it: $1/L$ (the inverse of the liquidity ratio).

Barometric firm price leadership Where the price leader is the one whose prices are believed to reflect market conditions in the most satisfactory way.

Barriers to entry Anything that prevents or impedes the entry of firms into an industry and thereby limits the amount of competition faced by existing firms.

Barter economy An economy where people exchange goods and services directly with one another without any payment of money. Workers would be paid with bundles of goods.

Base year (for index numbers) The year whose index number is set at 100.

Basic rate of tax The main marginal rate of tax, applying to most people's incomes.

Benefit drivers Attributes of a good or service that affect the benefit or utility from consuming it.

Benefits in kind Goods or services which the state provides directly to the recipient at no charge or at a subsidised price. Alternatively, the state can subsidise the private sector to provide them.

Bilateral monopoly Where a monopsony buyer faces a monopoly seller.

Bill of exchange A certificate promising to repay a stated amount on a certain date, typically three months from the issue of the bill. Bills pay no interest as such, but are sold at a discount and redeemed at face value, thereby earning a rate of discount for the purchaser.

Bounded rationality When the ability to make rational decisions is limited by lack of information or the time necessary to obtain such information.

Bretton Woods system An adjustable peg system whereby currencies were pegged to the US dollar. The USA maintained convertibility of the dollar into gold at the rate of $35 to an ounce.

Broad definitions of money Items in narrow definitions plus other items that can be readily converted into cash.

Broad money in UK (M4) Cash in circulation plus retail and wholesale bank and building society deposits.

Budget deficit The excess of an organisation's spending over its revenues. When applied to government it the excess of its spending over its tax receipts.

Budget surplus The excess of an organisation's revenues over its expenditures. When applied to government it the excess of its tax receipts over its spending.

Business cycle or trade cycle The periodic fluctuations of national output round its long-term trend.

Capital All inputs into production that have themselves been produced: e.g. factories, machines and tools.

Capital account of the balance of payments The record of the transfers of capital to and from abroad.

Capital adequacy ratio (CAR) The ratio of a bank's capital (reserves and shares) to its risk-weighted assets.

Capital accumulation An increase in the amount of capital that an economy has for production.

Capital deepening (shallowing) An increase (a decrease) in the amount of capital per worker (K/L).

Capital expenditure Investment expenditure; expenditure on assets.

Capital intensity The amount of physical capital that workers have to operate with and which can be measured by the amount of capital per worker (K/L).

Capital widening A proportionate rise in both capital and labour. There is more capital but no capital deepening.

Cartel A formal collusive agreement.

Central bank A country's central bank is banker to the government and the banks as a whole. In most countries the central bank operates monetary policy by setting interest rates and influencing the supply of money. The central bank in the UK is the Bank of England; in the eurozone it is the European Central Bank (ECB) and in the USA it is the Federal Reserve Bank (the 'Fed').

Centrally planned or command economy An economy where all economic decisions are taken by the central authorities.

Certificates of deposit (CDs) Certificates issued by banks for fixed-term interest-bearing deposits. They can be resold by the owner to another party.

Ceteris paribus Latin for 'other things being equal'. This assumption has to be made when making deductions from theories.

Change in demand This is the term used for a shift in the demand curve. It occurs when a determinant of demand *other* than price changes.

Change in supply The term used for a shift in the supply curve. It occurs when a determinant *other* than price changes.

Change in the quantity demanded The term used for a movement along the demand curve to a new point. It occurs when there is a change in price.

Change in the quantity supplied The term used for a movement along the supply curve to a new point. It occurs when there is a change in price.

Claimant unemployment Those in receipt of unemployment-related benefits.

Clearing system A system whereby inter-bank debts are settled.

Closed shop Where a firm agrees to employ only union members.

Coase theorem By sufferers from externalities doing deals with perpetrators (by levying charges or offering payments), the externality will be 'internalised' and the socially efficient level of output will be achieved.

Collateralised debt obligations (CDOs) These are a type of security consisting of a bundle of fixed-income assets, such as corporate bonds, mortgage debt and credit-card debt.

Collusive oligopoly Where oligopolists agree (formally or informally) to limit competition between themselves. They may set output quotas, fix prices, limit product promotion or development, or agree not to 'poach' each other's markets.

Collusive tendering Where two or more firms secretly agree on the prices they will tender for a contract. These prices will be above those which would be put in under a genuinely competitive tendering process.

Command-and-control (CAC) systems The use of laws or regulations backed up by inspections and penalties (such as fines) for non-compliance.

Commercial bills Bills of exchange issued by firms.

Common market A customs union where the member countries act as a single market with free movement of labour and capital, common taxes and common trade laws.

Comparative advantage A country has a comparative advantage over another in the production of a good if it can produce it at a lower opportunity cost: i.e. if it has to forgo less of other goods in order to produce it.

Competition for corporate control The competition for the control of companies through takeovers.

Complementary goods A pair of goods consumed together. As the price of one goes up, the demand for both goods will fall.

Compounding The process of adding interest each year to an initial capital sum.

Compromise strategy One whose worst outcome is better than the maximax strategy and whose best outcome is better than the maximin strategy.

Conglomerate merger When two firms in different industries merge.

Constrained discretion A set of principles or rules within which economic policy operates. These can be informal or enshrined in law.

Consumer durable A consumer good that lasts a period of time, during which the consumer can continue gaining utility from it.

Consumer sovereignty A situation where firms respond to changes in consumer demand without being in a position in the long run to charge a price above average cost.

Consumer surplus The excess of what a person would have been prepared to pay for a good (i.e. the utility) over what that person actually pays.

Consumption The act of using goods and services to satisfy wants. This will normally involve purchasing the goods and services.

Consumption function The relationship between consumption and national income. It can be expressed algebraically or graphically.

Consumption of domestically produced goods and services (C_d) The direct flow of money payments from households to firms.

Consumption smoothing The act by households of smoothing their levels of consumption over time despite facing volatile incomes.

Continuous market clearing The assumption that all markets in the economy continuously clear so that the economy is permanently in equilibrium.

Convergence of economies When countries achieve similar levels of growth, inflation, budget deficits as a percentage of GDP, balance of payments, etc.

Convergence in GDP per head The tendency for less rich developed countries to catch up with richer ones. Convergence does not apply to many of the poorer developing countries, however; the gap between them and richer countries has tended to widen.

Co-ordination failure When a group of firms (e.g. banks) acting independently could have achieved a more desirable outcome if they had co-ordinated their decision making.

Core workers Workers, normally with specific skills, who are employed on a permanent or long-term basis.

Cost–benefit analysis The identification, measurement and weighing up of the costs and benefits of a project in order to decide whether or not it should go ahead.

Cost-plus pricing (full-cost pricing) When firms price their product by adding a certain profit 'mark-up' to average cost.

Cost-push inflation Inflation caused by persistent rises in costs of production (independently of demand).

Countervailing power When the power of a monopolistic/oligopolistic seller is offset by powerful buyers who can prevent the price from being pushed up.

Cournot model A model of duopoly where each firm makes its price and output decisions on the assumption that its rival will produce a particular quantity.

Credibility of monetary policy The extent to which the public believes that the central bank will take the measures necessary to achieve the stated targets of monetary policy, for example an inflation rate target.

Credible threat (or promise) One that is believable to rivals because it is in the threatener's interests to carry it out.

Credit-constrained households Households which are limited in their ability to borrow against expected future incomes.

Credit cycle The expansion and contraction of credit flows over time.

Cross-price elasticity of demand The percentage (or proportionate) change in quantity demanded of one good divided by the percentage (or proportionate) change in the price of another.

Cross-price elasticity of demand (arc formula) ΔQ_{Da}/average $Q_{Da} \div \Delta P_b$/average P_b.

Cross-subsidise To use profits in one market to subsidise prices in another.

Crowding out Where increased public expenditure diverts money or resources away from the private sector.

Currency union A group of countries (or regions) using a common currency.

Current account balance of payments Exports of goods and services minus imports of goods and services plus net incomes and current transfers from abroad. If inflows of money (from the sale of exports, etc.) exceed outflows of money (from the purchase of imports, etc.), there is a 'current account surplus' (a positive figure). If outflows exceed inflows, there is a 'current account deficit' (a negative figure).

Current budget balance The difference between public-sector receipts and those expenditures classified as current rather than capital expenditures.

Current expenditure Recurrent spending on goods and factor payments.

Customs union A free-trade area with common external tariffs and quotas.

Cyclical or demand-deficient unemployment Disequilibrium unemployment caused by a fall in aggregate demand with no corresponding fall in the real wage rate.

Deadweight loss of an indirect tax The loss of consumer plus producer surplus from the imposition of an indirect tax.

Deadweight loss The loss in welfare (social surplus) arising from a socially inefficient allocation of resources.

Deadweight welfare loss The loss of consumer plus producer surplus in imperfect markets (when compared with perfect competition).

Debit card A card that has the same use as a cheque. Its use directly debits the person's current account.

Debt-servicing costs The costs incurred when repaying debt, including debt interest payments.

Decision tree (or game tree) A diagram showing the sequence of possible decisions by competitor firms and the outcome of each combination of decisions.

Deficit bias The tendency for frequent fiscal deficits and rising debt-to-GDP ratios because of the reluctance of policymakers to tighten fiscal policy

Deflationary gap The shortfall of national expenditure below national income (and injections below withdrawals) at the full-employment level of national income.

Deflationary policy Fiscal or monetary policy designed to reduce the rate of growth of aggregate demand.

Delegation of monetary policy The handing over by government of the operation of monetary policy to central banks.

Demand curve A graph showing the relationship between the price of a good and the quantity of the good demanded over a given time period. Price is measured on the vertical axis; quantity demanded is measured on the horizontal axis. A demand curve can be for an individual consumer or group of consumers, or more usually for the whole market.

Demand-deficient or cyclical unemployment Disequilibrium unemployment caused by a fall in aggregate demand with no corresponding fall in the real wage rate.

Demand management policies Demand-side policies (fiscal and/or monetary) designed to smooth out the fluctuations in the business cycle.

Demand-pull inflation Inflation caused by persistent rises in aggregate demand.

Demand schedule for an individual A table showing the different quantities of a good that a person is willing and able to buy at various prices over a given period of time.

Demand-side policies Policies designed to affect aggregate demand, and thereby the level of output, employment and prices. The two broad types of demand-side policy are fiscal policy and monetary policy.

Demand schedule (market) A table showing the different total quantities of a good that consumers are willing and able to buy at various prices over a given period of time.

Dependency Where the development of a developing country is hampered by its relationships with the industrialised world.

Depreciation (of a currency) A fall in the free-market exchange rate of the domestic currency with foreign currencies.

Depreciation (of capital) The decline in value of capital equipment due to age, or wear and tear.

Deregulation Where the government removes official barriers to competition (e.g. licences and minimum quality standards).

Derived demand The demand for a factor of production depends on the demand for the good that uses it.

Destabilising speculation Where the actions of speculators tend to make price movements larger.

Devaluation Where the government re-pegs the exchange rate at a lower level.

Diminishing marginal utility As more units of a good are consumed, additional units will provide less additional satisfaction than previous units.

Diminishing marginal utility of income Where each additional unit of income earned yields less additional utility than the previous unit.

Direct taxes Taxes on income and wealth. Paid directly to the tax authorities on that income or wealth.

Discount factor The value today of deciding to consume a good one period in the future as a proportion of the value when it is actually consumed.

Discount market A money market in which new or existing bills of exchange are bought and sold.

Discounting The process of reducing the value of future flows to give them a present valuation.

Discretionary fiscal policy Deliberate changes in tax rates or the level of government expenditure in order to influence the level of aggregate demand.

Diseconomies of scale Where costs per unit of output increase as the scale of production increases.

Disequilibrium unemployment Unemployment resulting from real wage rates in the economy being above the equilibrium level.

Disguised unemployment Where the same work could be done by fewer people.

Disposable income Original income plus cash benefits and minus direct taxes and other deductions.

Distribution of income by class of recipient Measurement of the distribution of income between the classes of person who receive it (e.g. homeowners and non-homeowners or those in the north and those in the south).

Distribution of income: functional distribution Measurement of the distribution of income according to the source of income (e.g. from employment, from profit, from rent, etc.).

Distribution of income: size distribution Measurement of the distribution of income according to the levels of income received by individuals (irrespective of source).

Distributive role for government intervention Interventions by government to affect the distribution of resources such as the distribution of incomes.

Diversification Where a firm expands into new types of business.

Divestiture aversion (or **endowment effect**) The hypothesis that people ascribe more value to things when they own them than when they are merely considering purchasing or acquiring them.

Domestically systemically important banks (D-SIBs) Banks identified by national regulators as being significant banks in the domestic financial system.

Dominant firm price leadership When firms (the followers) choose the same price as that set by a dominant firm in the industry (the leader).

Dominant strategy game Where the *same* policy is suggested by different strategies.

Dumping When exports are sold at prices below marginal cost – often as a result of government subsidy.

Duopoly An oligopoly where there are just two firms in the market.

Economic agents The general term for individuals, firms, government and organisations when taking part in economic activities such as buying, selling, saving, investing or in any other way interacting with other economic agents.

Economic efficiency A situation where each good is produced at the minimum cost and where individual people and firms get the maximum benefit from their resources.

Economic model The representation, either graphically, mathematically or in words, of the relationship between two or more variables. A model is a simplification of reality designed to explain just part of a complex process of economic relationships. It is thus based on various simplifying assumptions.

Economies of scale When increasing the scale of production leads to a lower cost per unit of output.

Economies of scope When increasing the range of products produced by a firm reduces the cost of producing each one.

ECU (European Currency Unit) The predecessor to the euro: a weighted average of EU currencies. It was used as a reserve currency and for the operation of the exchange rate mechanism (ERM).

Efficiency (allocative) A situation where the current combination of goods produced and sold gives the maximum satisfaction for each consumer at their current levels of income. Note that a redistribution of income would lead to a different combination of goods that was allocatively efficient.

Efficiency (productive) A situation where firms are producing the maximum output for a given amount of inputs, or producing a given output at the least cost.

Efficiency wage hypothesis The hypothesis that the productivity of workers is affected by the wage rate that they receive.

Efficiency wage rate The profit-maximising wage rate for the firm after taking into account the effects of wage rates on worker motivation, turnover and recruitment.

Elastic demand (with respect to price) Where quantity demanded changes by a larger percentage than price. Ignoring the negative sign, it will have a value greater than 1.

Elasticity A measure of the responsiveness of a variable (e.g. quantity demanded or quantity supplied) to a change in one of its determinants (e.g. price or income).

Endogenous growth theory A theory that the rate of economic growth depends on the rate of technological progress and diffusion, both of which depend on size of the capital stock and the capital goods industries, and also on institutions, incentives and the role of government.

Endogenous money supply Money supply that is determined (at least in part) by the demand for money.

Endowment effect (or **divestiture aversion**) The hypothesis that people ascribe more value to things when they own them than when they are merely considering purchasing or acquiring them.

Entrepreneurship The initiating and organising of the production of new goods, or the introduction of new techniques, and the risk-taking associated with it.

Envelope curve A long-run average cost curve drawn as the tangency points of a series of short-run average cost curves.

Environmental charges Charges for using natural resources (e.g. water or national parks), or for using the environment as a dump for waste (e.g. factory emissions or sewage).

Equation of exchange $MV = PY$. The total level of spending on GDP (MV) equals the total value of goods and services produced (PY) that go to make up GDP.

Equi-marginal principal (in consumption) Consumers will maximise total utility from their incomes by consuming that combination of goods where the ratio of the marginal utilities of goods is equal to the ratio of their prices.

Equi-marginal principle (in general). The optimum amount of two alternatives consumed (or produced) will be where the marginal benefit ratios of the two alternatives are equal to their marginal cost ratios.

Equilibrium A position of balance. A position from which there is no inherent tendency to move away.

Equilibrium ('natural') unemployment The difference between those who would like employment at the current wage rate and those willing and able to take a job.

Equilibrium price The price where the quantity demanded equals the quantity supplied: the price where there is no shortage or surplus.

Equities Company shares. Holders of equities are owners of the company and share in its profits by receiving dividends.

Equity A distribution of income that is considered to be fair or just. Note that an equitable distribution is not the same as an equal distribution and that different people have different views on what is equitable.

ERM (the exchange rate mechanism) A system of semi-fixed exchange rates used by most of the EU countries prior to adoption of the euro. Members' currencies were allowed to fluctuate against each other only within agreed bands. Collectively they floated against all other currencies.

Excess capacity (under monopolistic competition) In the long run, firms under monopolistic competition will produce at an output below their minimum-cost point.

Exchange equalisation account The gold and foreign exchange reserves account in the Bank of England.

Exchange rate The rate at which one national currency exchanges for another. The rate is expressed as the amount of one currency that is necessary to purchase *one unit* of another currency (e.g €1.15 = £1 and \$1.35 = £1).

Exchange rate band Where a currency is allowed to float between an upper and lower exchange rate, but is not allowed to move outside this band.

Exchange rate index or effective exchange rate A weighted average exchange rate expressed as an index where the value of the index is 100 in a given base year. The weights of the different currencies in the index add up to 1.

Exchange rate overshooting Where a fall (or rise) in the long-run equilibrium exchange rate causes the actual exchange rate to fall (or rise) by a greater amount before eventually moving back to the new long-run equilibrium level.

Exchange rate regime The system under which the government allows the exchange rate to be determined.

Exchange-rate transmission mechanism How a change in money supply affects aggregate demand via a change in exchange rates.

Exogenous money supply Money supply that does not depend on the demand for money but is set by the authorities.

Exogenous variable A variable whose value is determined independently of the model of which it is part.

Expectations-augmented Phillips curve A (short-run) Phillips curve whose position depends on the expected rate of inflation.

Expected value The predicted or average value of an outcome over a number of occurrences, calculated by taking each of the possible outcomes and multiplying it by its probability of occurrence and then adding each of these values.

Exponential discounting A method of reducing future benefits and costs to a present value. The discount rate depends on just how much less, from the consumer's perspective, *future* utility and costs (from a decision made today) are than gaining the utility/incurring the costs *today*.

Explicit costs The payments to outside suppliers of inputs.

External benefits Benefits from production (or consumption) experienced by people *other* than the producer (or consumer).

External costs Costs of production (or consumption) borne by people *other* than the producer (or consumer).

External diseconomies of scale Where a firm's costs per unit of output increase as the size of the whole industry increases.

External economies of scale Where a firm's costs per unit of output decrease as the size of the whole *industry* grows.

Externalities Costs or benefits of production or consumption experienced by society but not by the producers or consumers themselves. Sometimes referred to as 'spillover' or 'third-party' costs or benefits.

Factors of production (or resources) The inputs into the production of goods and services: labour, land and raw materials, and capital.

Final expenditure Expenditure on goods and services. This is included in GDP and is part of aggregate demand.

Final income Original income plus the addition of all benefits (cash and in kind) and the deduction of all taxes (direct and indirect).

Financial accelerator When a change in national income is amplified by changes in the financial sector, such as changes in interest rate differentials or the willingness of banks to lend.

Financial account of the balance of payments The record of the flows of money into and out of the country for the purposes of investment or as deposits in banks and other financial institutions.

Financial crowding out When an increase in government borrowing diverts money away from the private sector.

Financial deregulation The removal of or reduction in legal rules and regulations governing the activities of financial institutions.

Financial flexibility Where employers can vary their wage costs by changing the composition of their workforce or the terms on which workers are employed.

Financial instability hypothesis The theory that the economy goes through stages of credit accumulation which initially fuel aggregate demand but because of increasing financial distress eventually see a collapse in aggregate demand.

Financial instruments Financial products resulting in a financial claim by one party over another.

Financial intermediaries The general name for financial institutions (banks, building societies, etc.) which act as a means of channelling funds from depositors to borrowers.

Financialisation A term used to describe the process by which financial markets, institutions and instruments become increasingly significant in economies

Fine-tuning The use of demand management policy (fiscal or monetary) to smooth out cyclical fluctuations in the economy.

First-degree price discrimination Where a firm charges each consumer the maximum price they willing to pay for each unit.

First-mover advantage When a firm gains from being the first one to take action.

Fiscal drag The tendency of automatic fiscal stabilisers to reduce the recovery of an economy from recession.

Fiscal impulse A measure of the change in the fiscal stance arising from discretionary fiscal policy changes.

Fiscal policy Policy to affect aggregate demand by altering the balance between government expenditure and taxation.

Fiscal stance How deflationary or reflationary the Budget is.

Fixed costs Total costs that do not vary with the amount of output produced.

Fixed exchange rate (totally) Where the government takes whatever measures are necessary to maintain the exchange rate at some stated level.

Fixed factor An input that cannot be increased in supply within a given time period.

Flexible firm A firm that has the flexibility to respond to changing market conditions by changing the composition of its workforce.

Floating exchange rate When the government does not intervene in the foreign exchange markets, but simply allows the exchange rate to be freely determined by demand and supply.

Flow-of-funds equation The various items making up an increase (or decrease) in money supply

Forward exchange market Where contracts are made today for the price at which currency will be exchanged at some specified future date.

Framing The way in which a choice is presented or understood. A person may make different decisions depending on whether a choice is presented optimistically or pessimistically.

Franchising Where a firm is granted the licence to operate a given part of an industry for a specified length of time.

Free-market economy An economy where all economic decisions are taken by individual households and firms and with no government intervention.

Free-rider problem When it is not possible to exclude other people from consuming a good that someone has bought.

Free-trade area A group of countries with no trade barriers between them.

Freely floating exchange rate Where the exchange rate is determined entirely by the forces of demand and supply in the foreign exchange market with no government intervention whatsoever.

Frictional (search) unemployment Unemployment that occurs as a result of imperfect information in the labour market. It often takes time for workers to find jobs (even though there are vacancies) and in the meantime they are unemployed.

Full-employment level of national income The level of national income at which there is no deficiency of demand.

Functional distribution of income Measurement of the distribution of income according to the source of income (e.g. from employment, from profit, from rent, etc.).

Functional flexibility Where employers can switch workers from job to job as requirements change.

Functional separation of banks The ringfencing by banks of core retail banking services, such as deposit-taking, from riskier investment banking units.

Funding Where the authorities alter the balance of bills and bonds for any given level of government borrowing.

Future price A price agreed today at which an item (e.g. commodities) will be exchanged at some set date in the future.

Futures or forward market A market in which contracts are made to buy or sell at some future date at a price agreed today.

Gaia philosophy The respect for the rights of the environment to remain unharmed by human activity. Humans should live in harmony with the planet and other species. We have a duty to be

stewards of the natural environment, so that it can continue to be a self-maintaining and self-regulating system.

Game theory (or the theory of games) The study of alternative strategies oligopolists may choose to adopt, depending on their assumptions about their rivals' behaviour.

GDP (gross domestic product at market prices) The value of output (or income or expenditure) in terms of the prices actually paid. GDP = GVA + taxes on products − subsidies on products – subsidies on products.

GDP deflator The price index of all final domestically produced goods and services (domestic product): i.e. all items that contribute towards GDP.

General equilibrium A situation where all the millions of markets throughout the economy are in a simultaneous state of equilibrium.

General government debt The combined accumulated debt of central and local government.

General government deficit (or surplus) The combined deficit (or surplus) of central and local government.

Geographical immobility The lack of ability or willingness of people to move to jobs in other parts of the country.

Gini coefficient The area between the Lorenz curve and the 45° line divided by the total area under the 45° line.

Global systematically important banks (G-SIBs) Banks identified by a series of indicators as being significant banks in the global financial system.

GNI (see GNY)

GNY (gross national income) GDP plus net income from abroad.

Goodhart's Law Controlling a symptom of a problem, or only part of the problem, will not cure the problem: it will simply mean that the part that is being controlled now becomes a poor indicator of the problem.

Goods in short supply These are two goods where the production of more of one leads to the production of more of the other.

Government bonds or 'gilt-edged securities' A government security paying a fixed sum of money each year. It is redeemed by the government on its maturity date at its face value.

Government surplus (from a tax on a good) The total tax revenue earned by the government from sales of a good.

Grandfathering Where the number of emission permits allocated to a firm is based on its current levels of emission (e.g. permitted levels for all firms could be 80 per cent of their current levels).

Green tax A tax on output designed to charge for the adverse effects of production on the environment. The socially efficient level of a green tax is equal to the marginal environmental cost of production.

Gross domestic final expenditure Total expenditure by a country's residents on final goods and services. It thus includes expenditure on imports and excludes expenditure on exports

Gross domestic product (GDP) The value of output produced within the country over a 12-month period.

Gross income Original income plus cash benefits.

Gross national income (GNY) GDP plus net income from abroad.

Gross value added at basic prices (GVA) The sum of all the values added by all industries in the economy over a year. The figures exclude taxes on products (such as VAT) and include subsidies on products.

Heuristics People's use of strategies that draw on simple lessons from past experience when they are faced with similar, although not identical, choices.

Historic costs The original amount a firm paid for factors it now owns.

Hit-and-run competition When a firm enters an industry to take advantage of temporarily high profits and then leaves again as soon as the high profits have been exhausted.

Horizontal integration A business growth strategy that involves expanding within an existing market at the same stage of production by moving into allied products. An example would be an electricity supplier moving into gas supply or a car manufacturer moving into the production of coaches or heavy goods vehicles.

Horizontal merger When two firms in the same industry at the same stage in the production process merge.

Households' disposable income The income available for households to spend: i.e. personal incomes after deducting taxes on incomes and adding benefits.

Human capital The qualifications, skills and expertise that contribute to a worker's productivity.

Human Development Index (HDI) A composite index made up of three elements: an index for life expectancy, an index for school enrolment and adult literacy, and an index for GDP per capita (in PPP$).

Hysteresis The persistence of an effect even when the initial cause has ceased to operate. In economics, it refers to the persistence of unemployment even when the demand deficiency that caused it no longer exists.

Idle balances Money held for speculative purposes: money held in anticipation of a fall in asset prices.

Imperfect competition The collective name for monopolistic competition and oligopoly.

Implicit costs Costs that do not involve a direct payment of money to a third party, but which nevertheless involve a sacrifice of some alternative.

Import-substituting industrialisation (ISI) A strategy of restricting imports of manufactured goods and using the foreign exchange saved to build up domestic substitute industries.

Income effect (of a price change) The effect of a change in price on quantity demanded arising from the consumer becoming better or worse off as a result of the price change.

Income effect of a rise in wage rates Workers get a higher income for a given number of hours worked and may thus feel they need to work *fewer* hours as wage rates rise.

Income effect of a tax rise Tax increases reduce people's incomes and thus encourage people to work more.

Income elasticity of demand The percentage (or proportionate) change in quantity demanded divided by the percentage (or proportionate) change in income.

Income elasticity of demand (arc formula) $\Delta Q_D/\text{average } Q_D \div \Delta Y/\text{average } Y$.

Increasing opportunity costs of production When additional production of one good involves ever-increasing sacrifices of another.

Independence (of firms in a market) Where the decisions of one firm in a market will not have any significant effect on the demand curves of its rivals.

Independent risks Where two risky events are unconnected. The occurrence of one will not affect the likelihood of the occurrence of the other.

Index number The value of a variable expressed as 100 plus or minus its percentage deviation from a base year.

Indirect taxes Taxes on expenditure (e.g. VAT). Paid to the tax authorities, not by the consumer, but indirectly by the suppliers of the goods or services.

Indivisibilities The impossibility of dividing a factor into smaller units.

Induced investment Investment that firms make to enable them to meet extra consumer demand.

Industrial policies Policies to encourage industrial investment and greater industrial efficiency.

Industry's infrastructure The network of supply agents, communications, skills, training facilities, distribution channels, specialised financial services, etc. that supports a particular industry.

Inelastic demand Where quantity demanded changes by a smaller percentage than price. Ignoring the negative sign, it will have a value less than 1.

Infant industry An industry that has a potential comparative advantage, but which is as yet too underdeveloped to be able to realise this potential.

Inferior goods Goods whose demand *decreases* as consumer incomes increase. Such goods have a negative income elasticity of demand.

Inflation A general rise in the level of prices throughout the economy.

Inflation bias Excessive inflation that results from people raising their expectations of the inflation rate following expansionary demand management policy, encouraging government to loosen policy even further.

Inflation rate (annual) The percentage increase in prices over a 12-month period.

Inflationary gap The excess of national expenditure over income (and injections over withdrawals) at the full-employment level of national income.

Infrastructure (industry's) The network of supply agents, communications, skills, training facilities, distribution channels, specialised financial services, etc. that supports a particular industry.

Injections (*J*) Expenditure on the production of domestic firms coming from outside the inner flow of the circular flow of income. Injections equal investment (I_d) plus government expenditure (G_d) plus expenditure on exports (X).

Input–output analysis This involves dividing the economy into sectors where each sector is a user of inputs from and a supplier of outputs to other sectors. The technique examines how these inputs and outputs can be matched to the total resources available in the economy.

Insiders Those in employment who can use their privileged position (either as members of unions or because of specific skills) to secure pay rises despite an excess supply of labour (unemployment).

Integration: horizontal A business growth strategy that involves expanding within an existing market at the same stage of production by moving into allied products. An example would be an electricity supplier moving into gas supply or a car manufacturer moving into the production of coaches or heavy goods vehicles.

Integration: vertical A business growth strategy that involves expanding within an existing market, but at a different stage of production. Vertical integration can be 'forward', such as moving into distribution or retail, or 'backward', such as expanding into extracting raw materials or producing components.

Interdependence (under oligopoly) One of the two key features of oligopoly. Each firm will be affected by its rivals' decisions. Likewise its decisions will affect its rivals'. Firms recognise this interdependence. This recognition will affect their decisions.

Interest-rate transmission mechanism How a change in money supply affects aggregate demand via a change in interest rates.

International business cycle The cyclical nature of economic growth in the international economy, which both reflects the synchrony in national business cycles and also impacts on individual economies' cycles.

International harmonisation of economic policies Where countries attempt to co-ordinate their macroeconomic policies so as to achieve common goals.

International liquidity The supply of currencies in the world acceptable for financing international trade and investment.

International substitution effect As prices rise, people at home and abroad buy less of this country's products and more of products from abroad.

International trade multiplier The effect on national income in country B of a change in exports (or imports) of country A.

Intertemporal choices Choices where the costs and/or benefits occur over several periods and where therefore choices made at one moment in time affect future options.

Inter-temporal substitution effect Higher prices may lead to higher interest rates and thus less borrowing and more saving.

Intervention price (in the CAP) The price at which the EU is prepared to buy a foodstuff if the market price were to be below it.

Interventionist supply-side policies Policies to increase aggregate supply by government intervention to counteract the deficiencies of the market.

Investment The production of items that are not for immediate consumption.

Irrational exuberance Where banks and other economic agents are over confident about the economy and/or financial markets and expect economic growth to remain stronger and/or asset prices to rise further than warranted by evidence. The term is associated with the economist Robert Shiller and his book *Irrational Exuberance* (2000) and with the former US Federal Reserve Chairman, Alan Greenspan.

Joint float Where a group of currencies pegged to each other jointly float against other currencies.

Joint supply Where the production of more of one good leads to the production of more of another.

Keynesian consumption function The relationship between consumption and income. It can be expressed algebraically or graphically.

Kinked demand theory The theory that oligopolists face a demand curve that is kinked at the current price, demand being significantly more elastic above the current price than below. The effect of this is to create a situation of price stability.

Labour All forms of human input, both physical and mental, into current production.

Labour force The number employed plus the number unemployed.

Labour productivity Output per unit of labour, for example output per worker, and which is sometimes referred to as productivity.

Land (and raw materials) Inputs into production that are provided by nature: e.g. unimproved land and mineral deposits in the ground.

Law of comparative advantage Trade can benefit all countries if they specialise in the goods in which they have a comparative advantage.

Law of demand The quantity of a good demanded per period of time will fall as price rises and will rise as price falls, other things being equal (*ceteris paribus*).

Law of diminishing (marginal) returns When one or more factors are held fixed, there will come a point beyond which the extra output from additional units of the variable factor will diminish.

Law of large numbers The larger the number of events of a particular type, the more predictable will be their average outcome.

Lender of last resort The role of the Bank of England as the guarantor of sufficient liquidity in the monetary system.

Liabilities Claims by others on an individual or institution; debts of that individual or institution.

Liquidity The ease with which an asset can be converted into cash without loss.

Liquidity preference The demand for holding assets in the form of money.

Liquidity ratio The proportion of a bank's total assets held in liquid form.

Liquidity trap When interest rates are at their floor and thus any further increases in money supply will not be spent but merely be held in idle balances as people wait for the economy to recover and/or interest rates to rise.

Lock-outs Union members are temporarily laid off until they are prepared to agree to the firm's conditions.

Long run The period of time long enough for *all* factors to be varied.

Long-run average cost curve A curve that shows how average cost varies with output on the assumption that *all* factors are variable. (It is assumed that the least-cost method of production will be chosen for each output.)

Long-run marginal cost The extra cost of producing one more unit of output assuming that all factors are variable. (It is assumed that the least-cost method of production will be chosen for this extra output.)

Long-run profit maximisation An alternative theory of the firm which assumes that managers aim to shift cost and revenue curves so as to maximise profits over some longer time period.

Long-run shut-down point This is where the *AR* curve is tangential to the *LRAC* curve. The firm can just make normal profits. Any fall in revenue below this level will cause a profit-maximising firm to shut down once all costs have become variable.

Long run under perfect competition The period of time that is long enough for new firms to enter the industry.

Lorenz curve A curve showing the proportion of national income earned by any given percentage of the population (measured from the poorest upwards).

Lump-sum tax A tax of a fixed amount independent of the circumstances of the entity being taxed (e.g. independent of the output of a firm or the income of an individual).

Macroeconomic role for government intervention Interventions by government either to stabilise the economy in the short term or to promote longer-term economic growth.

Macroeconomics The branch of economics that studies economic aggregates (grand totals): e.g. the overall level of prices, output and employment in the economy.

Macro-prudential regulation Regulation which focuses on the financial system as a whole and which monitors its impact on the wider economy.

Managed floating A system of flexible exchange rates but where the government intervenes to prevent excessive fluctuations or even to achieve an unofficial target exchange rate.

Marginal benefit The additional benefit of doing a little bit more (or 1 unit more if a unit can be measured) of an activity.

Marginal capital-output ratio The amount of extra capital (in money terms) required to produce a £1 increase in national output. Since $I_i = \Delta K$, the marginal capital/output ratio $\Delta K/\Delta Y$ equals the accelerator coefficient (α).

Marginal consumer surplus The excess of utility from the consumption of one more unit of a good (*MU*) over the price paid: $MCS = MU - P$.

Marginal cost (of an activity) The additional cost of doing a little bit more (or 1 unit more if a unit can be measured) of an activity.

Marginal cost (of production) The cost of producing one more unit of output: $MC = \Delta TC/\Delta Q$.

Marginal disutility of work The extra sacrifice/hardship to a worker of working an extra unit of time in any given time period (e.g. an extra hour per day).

Marginal physical product The extra output gained by the employment of one more unit of the variable factor: $MPP = \Delta TPP/\Delta Q_v$.

Marginal productivity theory The theory that the demand for a factor depends on its marginal revenue product.

Marginal propensity to consume The proportion of a rise in national income that goes on consumption: $mpc = \Delta C/\Delta Y$.

Marginal propensity to consume domestically produced goods (mpc_d) The proportion of a rise in national income that is spent on goods and services produced within the country: $\Delta C_d/\Delta Y$.

Marginal propensity to import The proportion of an increase in national income that is spent on imports: $mpm = \Delta M/\Delta Y$.

Marginal propensity to save The proportion of an increase in national income saved: $mps = \Delta S/\Delta Y$.

Marginal propensity to withdraw The proportion of an increase in national income that is withdrawn from the circular flow: $mpw = \Delta W/\Delta Y$, where $mpw = mps + mpt + mpm$.

Marginal rate of income tax The income tax rate. The rate paid on each *additional* pound earned: $\Delta T/\Delta Y$.

Marginal revenue The extra revenue gained by selling one more unit per time period: $MR = \Delta TR/\Delta Q$.

Marginal revenue product (of a factor) The extra revenue a firm earns from employing one more unit of a variable factor: $MRP_{factor} = MPP_{factor} + MR_{good}$.

Marginal tax propensity The proportion of an increase in national income paid in tax: $mpt = \Delta T/\Delta Y$.

Marginal utility The extra satisfaction gained from consuming one extra unit of a good within a given time period.

Market demand schedule A table showing the different total quantities of a good that consumers are willing and able to buy at various prices over a given period of time.

Mark-up pricing or **average cost pricing** Where firms set the price by adding a profit mark-up to average cost.

Market The interaction between buyers and sellers.

Market clearing A market clears when supply matches demand, leaving no shortage or surplus.

Market for loanable funds The market for loans from and deposits into the banking system.

Market loans Short-term loans (e.g. money at call and short notice).

Market-orientated supply-side policies Policies to increase aggregate supply by freeing up the market.

Mark-up A profit margin added to average cost to arrive at price.

Mark-up pricing (or Average cost pricing) Where firms set the price by adding a profit mark-up to average cost.

Maturity gap The difference in the average maturity of loans and deposits.

Maturity transformation The transformation of deposits into loans of a longer maturity.

Maximum price A price ceiling set by the government or some other agency. The price is not allowed to rise above this level (although it is allowed to fall below it).

Mean (or **arithmetic mean**) The sum of the values of each of the members of the sample divided by the total number in the sample.

Means-tested benefits Benefits whose amount depends on the recipient's income or assets.

Median The value of the middle member of the sample.

Medium of exchange Something that is acceptable in exchange for goods and services.

Menu costs of inflation The costs associated with having to adjust price lists or labels.

Merit goods Goods which the government feels that people will under-consume and which therefore ought to be subsidised or provided free.

Microeconomics The branch of economics that studies individual units: e.g. households, firms and industries. It studies the interrelationships between these units in determining the pattern of production and distribution of goods and services.

Minimum price A price floor set by the government or some other agency. The price is not allowed to fall below this level (although it is allowed to rise above it).

Minimum reserve ratio A minimum ratio of cash (or other specified liquid assets) to deposits (either total or selected) that the central bank requires banks to hold.

Minsky moment A turning point in a credit cycle where a period of easy credit and rising debt is replaced by one of tight credit and debt consolidation.

Mixed economy An economy where economic decisions are made partly by the government and partly through the market.

Mixed market economy A market economy where there is some government intervention.

Mobility of labour The willingness and ability of labour to move to another job.

Monetarists Those who attribute inflation solely to rises in money supply.

Monetary base Notes and coin outside the central bank.

Monetary financial institutions (MFIs) Deposit-taking financial institutions including banks, building societies and central banks.

Monetary policy Policy to affect aggregate demand by altering the supply or cost of money (rate of interest).

Money illusion When people believe that a money wage or price increase (i.e. a nominal increase) represents a *real* increase: in other words, they ignore or underestimate inflation.

Money market The market for short-term loans and deposits.

Money multiplier The number of times greater the expansion of money supply (M_s) is than the expansion of the monetary base (M_b) that caused it: $\Delta M_s/\Delta M_b$.

Monopolistic competition A market structure where, like perfect competition, there are many firms and freedom of entry into the industry, but where each firm produces a differentiated product and thus has some control over its price.

Monopoly A market structure where there is only one firm in the industry.

Monopsony A market with a single buyer or employer.

Moral hazard The temptation to take more risk when you know that someone else (e.g. insurers) will cover the risks.

Multiplier (injections multiplier) The number of times a rise in income exceeds the rise in injections that caused it: $k = \Delta Y/\Delta J$.

Multiplier effect An initial increase in aggregate demand of £*x*m leads to an eventual rise in national income that is greater than £*x*m.

Multiplier formula (injections multiplier) The formula for the multiplier is $k = 1/mpw$ or $1/(1 - mpc_d)$.

Narrow definitions of money Items of money that can be spent directly (cash and money in cheque-book/debit-card accounts).

Nash equilibrium The position resulting from everyone making their optimal decision based on their assumptions about their rivals' decisions. Without collusion, there is no incentive for any firm to move from this position.

National debt The accumulated budget deficits (less surpluses) over the years: the total amount of government borrowing.

National expenditure on domestic product (E) Aggregate demand in the Keynesian model: i.e. $C_d + J$.

Nationalised industries State-owned industries that produce goods or services that are sold in the market.

Natural level of output The level of output in monetarist analysis where the vertical long-run aggregate supply curve cuts the horizontal axis.

Natural level of unemployment The level of equilibrium unemployment in monetarist analysis measured as the difference between the (vertical) long-run gross labour supply curve (N) and the (vertical) long-run effective labour supply curve (AS_L).

Natural monopoly A situation where long-run average costs would be lower if an industry were under monopoly than if it were shared between two or more competitors.

Natural rate of unemployment The rate of unemployment at which there is no excess or deficiency of demand for labour.

Natural rate hypothesis The theory that, following fluctuations in aggregate demand, unemployment will return to a natural rate. This rate is determined by supply-side factors, such as labour mobility.

Non-accelerating-inflation rate of unemployment (NAIRU The rate of unemployment consistent with a constant rate of inflation: the rate of unemployment at which the vertical long-run Phillips curve cuts the horizontal axis.

Natural wastage When a firm wishing to reduce its workforce does so by not replacing those who leave or retire.

Near money Highly liquid assets (other than cash).

Negative income tax A combined system of tax and benefits. As people earn more, they gradually lose their benefits until beyond a certain level they begin paying taxes.

Neoclassical analysis The analysis of market economies where it is assumed that individuals and firms are self-interested rational maximisers.

Net errors and omissions A statistical adjustment to ensure that the two sides of the balance of payments account balance. It is necessary because of errors in compiling the statistics.

Net investment Total investment minus depreciation.

Net national product (NNY) GNY minus depreciation.

Net worth The market value of a sector's stock of financial and non-financial wealth.

Network economies or network externalities The benefits to consumers of having a network of other people using the same product or service.

New classical school The school of economists which believes that markets clear virtually instantaneously and that expectations are formed 'rationally'.

New Keynesians Economists who seek to explain the downward stickiness of real wages and the resulting persistence of unemployment.

Nominal GDP GDP measured in current prices. These figures take no account of inflation.

Nominal values Money values measured at *current* prices.

Non-accelerating-inflation rate of unemployment (NAIRU) The rate of unemployment consistent with a constant rate of inflation. (In monetarist analysis, this is the same as the natural rate of unemployment: the rate of unemployment at which the vertical long-run Phillips curve cuts the horizontal axis.)

Non-bank private sector Households and non-bank firms. In other words, everyone in the country other than banks and the government (central and local).

Non-collusive oligopoly Where oligopolists have no agreement between themselves, either formal, informal or tacit.

Non-excludability Where it is not possible to provide a good or service to one person without it thereby being available for others to enjoy.

Non-price competition Competition in terms of product promotion (advertising, packaging, etc.) or product development.

Non-rivalry Where the consumption of a good or service by one person will not prevent others from enjoying it.

Normal goods Goods whose demand increases as consumer incomes increase. They have a positive income elasticity of demand. Luxury goods will have a higher income elasticity of demand than more basic goods.

Normal profit The opportunity cost of being in business: the profit that could have been earned in the next best alternative business. It is counted as a cost of production.

Normal rate of return The rate of return (after taking risks into account) that could be earned elsewhere.

Normative statement A value judgement.

Nudge theory The theory that positive reinforcement or making the decision easy can persuade people to make a particular choice. They are 'nudged' into so doing.

Numerical flexibility Where employers can change the size of their workforce as their labour requirements change.

Occupational immobility The lack of ability or willingness of people to move to other jobs irrespective of location.

Oligopoly A market structure where there are few enough firms to enable barriers to be erected against the entry of new firms.

Oligopsony A market with just a few buyers or employers.

Open economy One that trades with and has financial dealings with other countries.

Open-market operations The sale (or purchase) by the authorities of government securities in the open market in order to reduce (or increase) money supply or influence interest rates.

Opportunity cost Cost measured in terms of the best alternative forgone.

Optimal currency area The optimal size of a currency area is the one that maximises the benefits from having a single currency relative to the costs. If the area were increased or decreased in size, the costs would rise relative to the benefits.

Original income Income before taxes and benefits

Output gap The difference between actual and potential output. When actual output exceeds potential output, the gap is positive. When actual output is less than potential output, the gap is negative.

Outsiders Those out of work or employed on a casual, part-time or short-term basis, who have little or no power to influence wages or employment.

Overhead costs Costs arising from the general running of an organisation, and only indirectly related to the level of output.

Paradox of debt (or paradox of deleveraging) The paradox that one individual can increase his or her net worth by selling assets, but if this is undertaken by a large number of people aggregate net worth declines because asset prices fall.

Participation rate The percentage of the working-age population that is part of the workforce.

Perfect competition A market structure where there are many firms; where there is freedom of entry into the industry; where all firms produce an identical product; and where all firms are price takers.

Perfectly contestable market A market where there is free and costless entry and exit.

Phillips curve A curve showing the relationship between (price) inflation and unemployment. The original Phillips curve plotted *wage* inflation against unemployment for the years 1861–1957.

Picketing When people on strike gather at the entrance to the firm and attempt to dissuade workers or delivery vehicles from entering.

Pigouvian tax (or **subsidy**) A tax (or subsidy) designed to 'internalise' an externality. The marginal rate of a Pigouvian tax (or subsidy) should be equal to the marginal external cost (or benefit).

Plant economies of scale Economies of scale that arise because of the large size of the factory.

Policy ineffectiveness proposition The conclusion drawn from new classical models that, when economic agents anticipate changes in economic policy, output and employment remain at their equilibrium (or natural) levels.

Political business cycle The theory that governments will engineer an economic contraction, designed to squeeze out inflation, followed by a pre-election boom.

Poll tax A lump-sum tax per head of the population. Since it is a fixed *amount,* it has a marginal rate of zero with respect to both income and wealth.

Pooling (or **spreading**) **risks** (**for an insurance company**) The more policies an insurance company issues and the more independent the risks from these policies are, the more predictable will be the number of claims.

Portfolio balance The balance of assets, according to their liquidity, that people choose to hold in their portfolios.

Positive statement A value-free statement that can be tested by an appeal to the facts.

Post-tax income Disposable income minus indirect taxes.

Potential growth The percentage annual increase in the capacity of the economy to produce.

Potential output The economically sustainable level output that could be produced in the economy: i.e. one that involves a 'normal' level of capacity utilisation and does not result in rising inflation.

Poverty trap Where poor people are discouraged from working or getting a better job because any extra income they earn will be largely taken away in taxes and lost benefits.

Precautionary or **buffer-stock saving** Saving in response to uncertainty, for example uncertainty of future income.

Predatory pricing Where a firm sets its prices below average cost in order to drive competitors out of business.

Preferential trading arrangements A trade agreement whereby trade between the signatories is freer than trade with the rest of the world.

Present bias Time-inconsistent behaviour whereby people give greater weight to present payoffs relative to future ones than would be predicted by standard discounting techniques.

Present value (**in consumption**) The value a person places today on a good that will not be consumed until some point in the future.

Price benchmark A price that is typically used. Firms, when raising prices, will usually raise them from one benchmark to another.

Price discrimination Where a firm sells the same or similar product at different prices and the difference in price cannot be fully accounted for by any differences in the costs of supply.

Price discrimination: first degree Where a firm charges each consumer the maximum price they willing to pay for each unit.

Price discrimination: second degree Where a firm charges customers different prices for the same (or similar) product depending on factors such as the quantity or time purchased.

Price discrimination: third degree Where a firm divides consumers into different groups based on some characteristic that is relatively easy to observe and acceptable to the consumer. The firm then charges a different price to consumers in different groups, but the same price to all the consumers within each group.

Price elasticity of demand ($\mathbf{P\epsilon_D}$) The percentage (or proportionate) change in quantity demanded divided by the percentage (or proportionate) change in price: $\%\Delta Q_D \div \%\Delta P$.

Price elasticity of demand (**arc formula**) ΔQ/average $Q \div \Delta P$/average P. The average in each case is the average between the two points being measured.

Price elasticity of supply The percentage (or proportionate) change in quantity supplied divided by the percentage (or proportionate) change in price: $\%\Delta Q_S \div \%\Delta P$.

Price elasticity of supply (**arc formula**) ΔQ_S/average $Q_S \div \Delta P$/average P.

Price gap regulation Where the regulator puts a ceiling on the amount by which a firm can raise its price.

Price maker A firm that can choose the price it charges; it faces a downward-sloping demand curve. If, however, it alters its price, this will affect the quantity sold: a fall in price will lead to more being sold; a higher price will lead to less.

Price mechanism The system in a market economy whereby changes in price in response to changes in demand and supply have the effect of making demand equal to supply.

Price taker A person or firm with no power to be able to influence the market price.

Primary labour market The market for permanent full-time core workers.

Principal–agent problem Where people (principals), as a result of lack of knowledge, cannot ensure that their best interests are served by their agents.

Principle of diminishing marginal utility As more units of a good are consumed, additional units will provide less additional satisfaction than previous units.

Prisoners' dilemma Where two or more firms (or people), by attempting independently to choose the best strategy for whatever the other(s) are likely to do, end up in a worse position than if they had co-operated in the first place.

Private efficiency Where a person's marginal benefit from a given activity equals the marginal cost.

Private limited company A company owned by its shareholders. Shareholders' liability is limited to the value of their shares. Shares can only be bought and sold privately.

Producer surplus The excess of total revenue over total cost: i.e. profit.

Product differentiation When one firm's product is sufficiently different from its rivals' to allow it to raise the price of the product without customers all switching to the rivals' products. A situation where a firm faces a downward-sloping demand curve.

Production The transformation of inputs into outputs by firms in order to earn profit (or meet some other objective).

Production function for the economy (or **aggregate production function**) The relationship between the economy's output and output per worker (labour productivity), holding the level of human capital and the state of technology constant.

Production possibility curve A curve showing all the possible combinations of two goods that a country can produce within a specified time period with all its resources fully and efficiently employed.

Productive efficiency A situation where firms are producing the maximum output for a given amount of inputs, or producing a given output at the least cost.

Productivity deal When, in return for a wage increase, a union agrees to changes in working practices that will increase output per worker.

Profit (rate of) Total profit ($T\Pi$) as a proportion of the total capital employed (K): $r = T\Pi/K$.

Profit-maximising rule Profit is maximised where marginal revenue equals marginal cost.

Profit satisficing Where decision-makers in a firm aim for a target level of profit rather than the absolute maximum level.

Progressive tax A tax whose average rate with respect to income rises as income rises.

Proportional tax A tax whose average rate with respect to income stays the same as income rises.

Prudential control The insistence by the Bank of England that banks maintain adequate liquidity.

Public good A good or service that has the features of non-rivalry and non-excludability and as a result would not be provided by the free market.

Public limited company A company owned by its shareholders. Shareholders' liability is limited to the value of their shares. Shares may be bought and sold publicly – on the stock exchange.

Public-sector borrowing requirement The old name for the public-sector net cash requirement.

Public-sector debt repayment (PSDR) or **Public-sector surplus** The old name for a negative public-sector net cash requirement. The (annual) surplus of the public sector, and thus the amount of debt that can be repaid.

Public-sector net borrowing (PSNB) The difference between the expenditures of the public sector and its receipts from taxation and the revenues from public corporations.

Public-sector net cash requirement (PSNCR) The (annual) deficit of the public sector, and thus the amount that the public sector must borrow.

Public-sector net debt Gross public-sector debt minus liquid financial assets.

Purchasing-power parity exchange rate The rate of exchange of a country's currency into the US dollar that would allow a given amount of that currency to buy the same amount of goods in the USA as within the country concerned.

Pure fiscal policy Fiscal policy that does not involve any change in money supply.

Quantitative easing A deliberate attempt by the central bank to increase the money supply by buying large quantities of securities through open-market operations. These securities could be securitised mortgage and other private-sector debt or government bonds. It uses electronic money (reserve liabilities) created specifically for this purpose.

Quantity demanded The amount of a good a consumer is willing and able to buy at a given price over a given period of time.

Quantity theory of money The price level (P) is directly related to the quantity of money in the economy (M).

Quota (set by a cartel) The output that a given member of a cartel is allowed to produce (production quota) or sell (sales quota).

Rate of economic growth The percentage increase in output between two moments of time, typically over a 12-month period.

Rate of profit Total profit ($T\Pi$) as a proportion of the capital employed (K): $r = T\Pi/K$.

Rational choices Choices that involve weighing up the benefit of any activity against its opportunity cost.

Rational consumer A person who weighs up the costs and benefits to him or her of each additional unit of a good purchased.

Rational consumer behaviour The attempt to maximise total consumer surplus: i.e. the attempt to get as much value as possible from your money when purchasing a good. If $MU > P$, you will buy more; if $MU < P$, you will buy less (or not buy at all); if $MU = P$, you will maintain your current level of consumption.

Rational economic behaviour Doing more of activities whose marginal benefit exceeds their marginal cost and doing less of those activities whose marginal cost exceeds their marginal benefit.

Rational expectations Expectations based on the *current* situation. These expectations are based on the information people have to hand. Whilst this information may be imperfect and therefore people will make errors, these errors will be random.

Rational firm behaviour When a firm weighs up the costs and benefits of alternative courses of action and then seeks to maximise its net benefit.

Rational producer behaviour When a firm weighs up the costs and benefits of alternative courses of action and then seeks to maximise its net benefit.

Rationalisation The reorganising of production (often after a merger) so as to cut out waste and duplication and generally to reduce costs.

Rationing Where the government restricts the amount of a good that people are allowed to buy.

Real balance effect As the price level rises, the value of people's money assets falls. They therefore spend less in their attempt to protect the real value of their savings.

Real business cycle theory The new classical theory that explains cyclical fluctuations in terms of shifts in aggregate supply, rather than aggregate demand.

Real GDP GDP measured in constant prices that ruled in a chosen base year, such as 2000 or 2015. These figures *do* take account of inflation. When inflation is positive, real GDP figures will grow more slowly than nominal GDP figures.

Real growth values Values of the rate of growth of GDP or any other variable after taking inflation into account. The real value of the growth in a variable equals its growth in money (or 'nominal') value minus the rate of inflation.

Real income Income measured in terms of how much it can buy. If your *money* income rises by 10 per cent, but prices rise by 8 per cent, you can only buy 2 per cent more goods than before. Your *real* income has risen by 2 per cent.

Real values Money values corrected for inflation.

Real-wage unemployment Disequilibrium unemployment caused by real wages being driven up above the market-clearing level.

Recession A period where national output falls for six months or more.

Recessionary or deflationary gap The shortfall of national expenditure below national income (and injections below withdrawals) at the full-employment level of national income.

Reciprocity (in economics) Where people's behaviour is influenced by the effects it will have on others.

Recognised banks Banks licensed by the Bank of England. All financial institutions using the word 'bank' in their title have to be recognised by the Bank of England. This requires them to have paid-up capital of at least £5 million and to meet other requirements about their asset structure and range of services.

Rediscounting bills of exchange Buying bills before they reach maturity.

Reference dependence The situation where people evaluate the outcomes of choices relative to a reference point and then classify them as either gains and losses.

Reference dependent loss aversion Where people classify outcomes as either losses or gains in relation to a reference point and the perceived losses are disliked more than would be predicted by standard diminishing marginal utility.

Reflationary policy Fiscal or monetary policy designed to increase the rate of growth of aggregate demand.

Regional multiplier effects When a change in injections into or withdrawals from a particular region causes a multiplied change in income in that region.

Regional unemployment Structural unemployment occurring in specific regions of the country.

Regressive tax A tax whose average rate with respect to income falls as income rises.

Regulatory capture Where the regulator is persuaded to operate in the industry's interests rather than those of the consumer.

Regulatory role for government intervention Interventions by government to regulate economic activity through legally enforceable rules or actions.

Relative price The price of one good compared with another (e.g. good X is twice the price of good Y).

Repos Sale and repurchase agreements. An agreement between two financial institutions whereby one in effect borrows from another by selling it assets, agreeing to buy them back (repurchase them) at a fixed price and on a fixed date.

Restrictive practice Where two or more firms agree to adopt common practices to restrict competition.

Retail banking Branch, telephone, postal and Internet banking for individuals and businesses at published rates of interest and charges. Retail banking involves the operation of extensive branch networks.

Retail deposits and loans Deposits and loans made through bank/building society branches at published interest rates.

Retail price index (RPI) An index of the prices of goods bought by a typical household.

Revaluation Where the government re-pegs the exchange rate at a higher level.

Reverse repos When gilts or other assets are *purchased* under a sale and repurchase agreement. They become an asset to the purchaser.

Rights issue An issue of additional shares confined to existing shareholders who have the right to buy them at a predetermined price.

Risk When an outcome may or may not occur, but its probability of occurring is known. It is a measure of the variability of that outcome.

Risk averse A person not prepared to take a gamble even if the odds are favourable.

Risk loving Where you would be willing to take a gamble even if its expected value was lower than that of certain outcome. The more risk loving you are, the lower the expected value you would be prepared to accept (i.e. the worse the odds would need to be).

Risk neutral When a person is indifferent between a certain outcome and a gamble with the same expected value.

Risk transformation The ability of financial institutions to spread risks by having a large number of clients.

Sale and repurchase agreement (repos) An agreement between two financial institutions whereby one in effect borrows from another by selling it assets, agreeing to buy them back (repurchase them) at a fixed price and on a fixed date.

Sales revenue maximisation An alternative theory which assumes that managers aim to maximise the firm's short-run total revenue.

Scarcity The excess of human wants over what can actually be produced to fulfil these wants.

Search theory This examines people's behaviour under conditions of ignorance where it takes time to search for information.

Seasonal unemployment Unemployment associated with industries or regions where the demand for labour is lower at certain times of the year.

Second-degree price discrimination Where a firm charges customers different prices for the same (or similar) product depending on factors such as the quantity or time purchased

Secondary action Industrial action taken against a company not directly involved in a dispute (e.g. a supplier of raw materials to a firm whose employees are on strike).

Secondary labour market The market for peripheral workers, usually employed on a temporary or part-time basis, or a less secure 'permanent' basis.

Secondary marketing Where assets are sold before maturity to another institution or individual.

Securitisation Where future cash flows (e.g. from interest rate or mortgage payments) are turned into marketable securities, such as bonds.

Self-fulfilling speculation The actions of speculators tend to cause the very effect that they had anticipated.

Sequential move games One firm (the first mover) makes and implements a decision. Rival firms (second moves) can observe the actions taken by the first mover before making their own decisions.

Set-aside A system in the EU of paying farmers not to use a certain proportion of their land.

Short run (in production) The period of time over which at least one factor is fixed.

Short-run shut-down point This is where the *AR* curve is tangential to the *AVC* curve. The firm can only just cover its variable costs. Any fall in revenue below this level will cause a profit-maximising firm to shut down immediately.

Short run under perfect competition The period during which there is too little time for new firms to enter the industry.

Short selling (or **shorting**) Where investors borrow an asset, such as shares or foreign currency; sell the asset, hoping the price will soon fall; then buy it back later and return it to the lender. Assuming the price has fallen, the short seller will make a profit of the difference (minus any fees).

Sight deposits Deposits that can be withdrawn on demand without penalty.

Simultaneous single-move game A game where each player has just one move, where each player plays at the same time and acts without knowledge of the actions chosen by other players.

Size distribution of income Measurement of the distribution of income according to the levels of income received by individuals (irrespective of source).

Social benefit Private benefit plus externalities in consumption.

Social capital (OECD definition) Networks, together with shared norms, values and understandings, that facilitate co-operation within of among groups.

Social cost Private cost plus externalities in production.

Social efficiency Production and consumption at the point where marginal social benefit equals marginal social cost ($MSB = MSC$).

Social mobility The ease with which people can move within or between social strata, such as those defined by wealth, income, class, occupation, etc.

Social surplus Total social benefits minus total social costs.

Socially optimal output The output where $MSC = MSB$: the output where total social surplus is maximised.

Solow growth model A model which explains economic growth in terms of the effects on the capital stock and output of a change in investment.

Special deposits A system used up to 1980. Deposits that the banks could be required to make in the Bank of England. They remained frozen there until the Bank of England chose to release them.

Special purpose vehicle (SPV) Legal entities created by financial institutions for conducting specific financial functions, such as bundling assets together into fixed interest bonds and selling them.

Specialisation and division of labour Where production is broken down into a number of simpler, more specialised tasks, thus allowing workers to acquire a high degree of efficiency.

Specific tax A tax on a good levied at a fixed amount per unit of the good, irrespective of the price of that unit.

Speculation Where people make buying or selling decisions based on their anticipations of future prices.

Speculators People who buy (or sell) commodities or financial assets with the intention of profiting by selling them (or buying them back) at a later date at a higher (lower) price.

Spot price The current market price.

Spreading (or pooling) risks (for an insurance company) The more policies an insurance company issues and the more independent the risks from these policies are, the more predictable will be the number of claims.

Stabilising speculation Where the actions of speculators tend to reduce price fluctuations.

Stakeholders (in a company) People who are affected by a company's activities and/or performance (customers, employees, owners, creditors, people living in the neighbourhood, etc.). They may or may not be in a position to take decisions, or influence decision taking, in the firm.

Standardised unemployment rate The measure of the unemployment rate used by the ILO and OECD. The unemployed are defined as persons of working age who are without work, available to start work within two weeks and either have actively looked for work in the last four weeks or are waiting to take up an appointment.

Status quo bias Where, other things being equal, people are averse to change, preferring things to stay as they are.

Steady-state real national income The long-run equilibrium level of real national income (real GDP) per worker. The level at which all investment is used merely to maintain the existing capital stock per worker at its current level.

Strategic trade theory The theory that protecting/supporting certain industries can enable them to compete more effectively with large monopolistic rivals abroad. The effect of the protection is to increase long-run competition and may enable the protected firms to exploit a comparative advantage that they could not have done otherwise.

Structural deficit (or surplus) The public-sector deficit (or surplus) that would occur if the economy were operating at the potential level of national income: i.e. one where there is a zero output gap.

Structural unemployment Unemployment that arises from changes in the pattern of demand or supply in the economy. People made redundant in one part of the economy cannot immediately take up jobs in other parts (even though there are vacancies).

Sub-prime debt Debt where there is a high risk of default by the borrower (e.g. mortgage holders who are on low incomes facing higher interest rates and falling house prices).

Substitute goods A pair of goods that are considered by consumers to be alternatives to each other. As the price of one goes up, the demand for the other rises.

Substitutes in supply These are two goods where an increased production of one means diverting resources away from producing the other.

Substitution effect of a price change The effect of a change in price on quantity demanded arising from the consumer switching to or from alternative (substitute) products.

Substitution effect of a rise in wage rates Workers will tend to substitute income for leisure as leisure now has a higher opportunity cost. This effect leads to *more* hours being worked as wage rates rise.

Substitution effect of a tax rise Tax increases reduce the opportunity cost of leisure and thus encourage people to work less.

Sunk cost fallacy The tendency for individuals to continue with activities because of the resources they have previously invested in them.

Sunk costs (in production) Costs that cannot be recouped (e.g. by transferring assets to other uses). Examples include specialised machinery or the costs of an advertising campaign.

Supernormal profit (also known as **pure profit**, **economic profit**, **abnormal profit**, or simply **profit**) The excess of total profit above normal profit.

Supply curve A graph showing the relationship between the price of a good and the quantity of the good supplied over a given period of time.

Supply schedule A table showing the different quantities of a good that producers are willing and able to supply at various prices over a given time period. A supply schedule can be for an individual producer or group of producers, or for all producers (the market supply schedule).

Supply-side economics An approach that focuses directly on aggregate supply and how to shift the aggregate supply curve outwards.

Supply-side policy Government policy that attempts to alter the level of aggregate supply directly (rather than through changes in aggregate demand). Such policies can be interventionist or market-orientated.

Sustainability (environmental) The ability of the environment to survive its use for economic activity.

Sustainable economic growth (environmental) Economic growth which sustains the planet's natural resources for future generations.

Sustainable output The level of national output corresponding to no excess or deficiency of aggregate demand.

Switching costs The costs to a consumer of switching to an alternative supplier.

Tacit collusion Where oligopolists take care not to engage in price cutting, excessive advertising or other forms of competition. There may be unwritten 'rules' of collusive behaviour such as price leadership.

Takeover bid Where one firm attempts to purchase another by offering to buy the shares of that company from its shareholders.

Tariff escalation The system whereby tariff rates increase the closer a product is to the finished stage of production.

Tariffs (or **import levies**) Taxes on imported products: i.e. customs duties.

Tax allowance An amount of income that can be earned tax-free. Tax allowances vary according to a person's circumstances.

Taylor rule A rule adopted by a central bank for setting the rate of interest. It will raise the interest rate if (a) inflation is above target or (b) real national income is above the sustainable level (or unemployment is below the equilibrium rate). The rule states how much interest rates will be changed in each case.

Technological unemployment Structural unemployment that occurs as a result of the introduction of labour-saving technology.

Terms of trade The price index of exports divided by the price index of imports and then expressed as a percentage. This means that the terms of trade will be 100 in the base year.

Third-degree price discrimination Where a firm divides consumers into different groups based on some characteristic that is relatively easy to observe and acceptable to the consumer. The firm then charges a different price to consumers in different groups, but the same price to all the consumers within each group.

Time consistency Where a person's preferences remain the same over time. For example, it is time consistent if you plan to buy a book when your student loan arrives and then actually do so when it does.

Time deposits Deposits that require notice of withdrawal or where a penalty is charged for withdrawals on demand.

Tit-for-tat strategy Where a firm will always respond to a price cut by a rival by doing the same.

Total consumer expenditure on a product (*TE*) (per period of time) The price of the product multiplied by the quantity purchased: $TE = P \times Q$.

Total consumer surplus The excess of a person's total utility from the consumption of a good (TU) over the amount that person spends on it (TE): $TCS = TU - TE$.

Total cost The sum of total fixed costs and total variable costs: $TC = TFC + TVC$.

Total physical product The total output of a product per period of time that is obtained from a given amount of inputs.

Total (private) surplus Total consumer surplus ($TU - TE$) plus total producer surplus ($TR - TVC$).

Total producer surplus (*TPS*) Total revenue minus total variable cost ($TR - TVC$): in other words, total profit plus total fixed cost ($T\Pi + TFC$).

Total revenue (*TR*) (per period of time) The total amount received by firms from the sale of a product, before the deduction of taxes or any other costs. The price multiplied by the quantity sold: $TR = P + Q$.

Total social surplus Total benefits to society from consuming a good minus total costs to society from producing it. In the absence of externalities, total social surplus is the same as total (private) surplus.

Total utility The total satisfaction a consumer gets from the consumption of all the units of a good consumed within a given time period.

Tradable permits Each firm is given a permit to produce a given level of pollution. If less than the permitted amount is produced, the firm is given a credit. This can then be sold to another firm, allowing it to exceed its original limit.

Trade creation Where a customs union leads to greater specialisation according to comparative advantage and thus a shift in production from higher-cost to lower-cost sources.

Trade cycle or business cycle The periodic fluctuations of national output round its long-term trend.

Trade diversion Where a customs union diverts consumption from goods produced at a lower cost outside the union to goods produced at a higher cost (but tariff free) within the union.

Traditional theory of the firm The analysis of pricing and output decisions of the firm under various market conditions, assuming that the firm wishes to maximise profit.

Transfer payments Moneys transferred from one person or group to another (e.g. from the government to individuals) without production taking place.

Transfers Transfers of money from taxpayers to recipients of benefits and subsidies. They are not an injection into the circular flow but are the equivalent of a negative tax (i.e. a negative withdrawal).

Transmission mechanism The process by which a change in a policy instrument (such as interest rates or taxation) affects economic outcomes (such as inflation or unemployment).

Treasury bills Bills of exchange issued by the Bank of England on behalf of the government. They are a means whereby the government raises short-term finance.

Trigger strategy Once a firm observes that its rival has broken some agreed behaviour, it will never co-operate with them again.

Uncertainty When an outcome may or may not occur and its probability of occurring is not known.

Underemployment When people work fewer hours than they would like at their current wage rate. *International Labour Organisation (ILO) definition*: a situation where people currently working less than 'full time' would like to work more hours (at current wage rates), either by working more hours in their current job, or by switching to an alternative job with more hours or by taking on an additional part-time job or any combination of the three. *Eurostat definition*: where people working less than 40 hours per week would like to work more hours in their current job at current wage rates.

Underground markets Where people ignore the government's price and/or quantity controls and sell illegally at whatever price equates illegal demand and supply.

Unemployed (economist's definition) Those of working age who are without work, but who are available for work at current wage rates.

Unemployment rate The number unemployed expressed as a percentage of the labour force.

Unit elastic demand Where quantity demanded changes by the same percentage as price. Ignoring the negative sign, it will have a value equal to 1.

Universal benefits Benefits paid to everyone in a certain category irrespective of their income or assets.

Value added tax (VAT) A tax on goods and services, charged at each stage of production as a percentage of the value added at that stage.

Variable costs Total costs that vary with the amount of output produced.

Variable factor An input that can be increased in supply within a given time period.

Velocity of circulation The number of times annually that money on average is spent on goods and services that make up GDP.

Vent for surplus When international trade enables a country to exploit resources that would otherwise be unused.

Vertical integration A business growth strategy that involves expanding within an existing market, but at a different stage of production. Vertical integration can be 'forward', such as moving into distribution or retail, or 'backward', such as expanding into extracting raw materials or producing components.

Wage–price spiral Wages and prices chasing each other as the aggregate demand curve continually shifts to the right and the aggregate supply curve continually shifts upwards.

Wage taker An employer (or employee) who is unable to influence the wage rate.

Weighted average The average of several items, where each item is ascribed a weight according to its importance. The weights must add up to 1.

Wholesale banks Banks specialising in large-scale deposits and loans and dealing mainly with companies. Interest rates and charges may be negotiable.

Wholesale deposits and loans Large-scale deposits and loans made by and to firms at negotiated interest rates.

Wide monetary base (MO) Notes and coin outside the central bank plus banks' operational deposits with the central bank.

Withdrawals (W) (or leakages) Incomes of households or firms that are not passed on round the inner flow. Withdrawals equal net saving (S) plus net taxes (T) plus expenditure on imports (M): $W = S + T + M$.

Working to rule Workers do the bare minimum they have to, as set out in their job descriptions.

Index

Page numbers in **bold** indicate definitions

Explore the essential principles of this exciting subject and engage with real-life issues facing individuals, businesses, society and economists today.

Essentials of Economics, 9th edition by John Sloman and Dean Garratt provides a clear, concise and engaging introduction to economics, making it the ideal textbook if you are studying on a one-semester or non-specialist course.

The new edition has been thoroughly updated to include insights into and analysis of the economic effects of real global problems such as the climate emergency, the Russian invasion of Ukraine, the COVID-19 pandemic and the cost of living crisis. It examines various government policies to address the economic consequences of these problems.

Each chapter includes topical examples, news stories and case studies to explain and illustrate key economic concepts as well as activities, questions and useful summaries to help you check your understanding.

Access the free student website that accompanies this book for additional learning support, including animated explainers of key economic models, 225 extra case studies and answers to in-text questions.

About the authors:

John Sloman was Director of the Economics Network from its foundation in 1999 until 2012, and is now Visiting Fellow at the University of Bristol where the Network is based. John is also Visiting Professor at the University of the West of England, Bristol.

Dean Garratt is a Senior Teaching Fellow at Aston Business School and Associate of the Economics Network. Dean was also an academic assessor for the Government Economic Service and a tutor on HM Treasury's Graduate Development Programme.

Also available with MyLab Economics:

MyLab® is the flexible teaching and learning platform that combines trusted author content with digital assessments and help tools to use at point-of-need. **MyLab Economics** for this book includes a comprehensive study guide with practice questions, quizzes and tests – all with immediate and automatic feedback to help support your learning.

Pearson, the world's *learning* company.

www.pearson.com

ISBN 978-1-29244-010-1

Pearson

Cover image © w3DProfi / Shutterstock
Cover design by Kelly Miller