



UK Edition

Making Everything Easier!™

2nd Edition

Economics

FOR

DUMMIES®

Learn to:

- Look through economic history and spot the trends
- Understand micro- and macroeconomics
- Get to grips with consumer behaviour and its influence on the economy
- Spot the signs of a recession and see how economic decisions affect you

Peter Antonioni

*Senior Teaching Fellow,
University College London*

Sean Masaki Flynn, PhD

Assistant Professor of Economics, Scripps College



Get More and Do More at Dummies.com®



Start with **FREE** Cheat Sheets

Cheat Sheets include

- Checklists
- Charts
- Common Instructions
- And Other Good Stuff!

To access the Cheat Sheet created specifically for this book, go to
www.dummies.com/cheatsheet/economicsuk

Get Smart at Dummies.com

Dummies.com makes your life easier with thousands of answers on everything from removing wallpaper to using the latest version of Windows.

Check out our

- Videos
- Illustrated Articles
- Step-by-Step Instructions

Want a weekly dose of Dummies?

Sign up for Newsletters on

- Digital Photography
- Microsoft Windows & Office
- Personal Finance & Investing
- Health & Wellness
- Computing, iPods & Mobile Phones
- eBay
- Internet
- Food, Home & Garden

Find out "HOW" at Dummies.com



Economics
FOR
DUMMIES®
2ND EDITION

by Peter Antonioni and Sean Masaki Flynn



A John Wiley and Sons, Ltd, Publication

Economics For Dummies[®], 2nd Edition

Published by
John Wiley & Sons, Ltd
The Atrium
Southern Gate
Chichester
West Sussex
PO19 8SQ
England

E-mail (for orders and customer service enquires): cs-books@wiley.co.uk

Visit our Home Page on www.wiley.com

Copyright © 2011 John Wiley & Sons, Ltd, Chichester, West Sussex, England

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London, W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (44) 1243 770620.

Trademarks: Wiley, the Wiley Publishing logo, For Dummies, the Dummies Man logo, A Reference for the Rest of Us!, The Dummies Way, Dummies Daily, The Fun and Easy Way, Dummies.com and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. Wiley Publishing, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER, THE AUTHOR, AND ANYONE ELSE INVOLVED IN PREPARING THIS WORK MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, please contact our Customer Care Department within the US at 877-762-2974, outside the US at 317-572-3993, or fax 317-572-4002.

For technical support, please visit www.wiley.com/techsupport.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

British Library Cataloguing in Publication Data: A catalogue record for this book is available from the British Library

ISBN: 978-0-470-97325-7 (paperback), 978-0-470-97369-1 (ebk), 978-0-470-97365-3 (ebk), 978-0-470-97366-0 (ebk)

Printed and bound in Great Britain by TJ International Ltd

10 9 8 7 6 5 4 3 2 1



About the Authors

Peter Antonioni was educated at Pembroke College, Oxford, and Birkbeck College, London, and has worked in both academic and private sectors as an economist before finally ending up as Lecturer in Management at University College London.

His research interests are diverse, and include the football transfer market, the process of enterprise in technology markets, and the techniques and practice of the arcane field of Long Range Scenario Planning.

His great passions include composing and recording electronic music, practicing the mystical martial arts of ancient China, and weeping with uncontrollable dismay at Spurs' latest setbacks. Amongst his talents are an uncanny ability to recall every plot point and line of dialogue from *Battlestar Galactica*, and the superpower of creating a mean *penne arrabiata*. He blogs some of the above at pja.typepad.com.

Sean Masaki Flynn earned his Ph.D. in Economics at the University of California, Berkeley, studying under Nobel Prize winners George Akerlof and Daniel McFadden.

He is a member of the American Economic Association, the American Finance Association, the Economic Science Association, and the Society for the Advancement of Behavioral Economics.

His research focuses on the often puzzling and seemingly irrational behavior of stock market investors, but he's also investigated topics as wide-ranging as the factors that affect customer tipping behaviour at restaurants and why you see a lot of unionised workers only in certain industries. He's also a leading expert on closed-end mutual funds.

His great passion is the Japanese martial art of aikido, which he has taught for over a decade to thousands of students both in the United States and abroad. If you like the martial arts, you might enjoy reading his book, *Shodokan Aikido: Basics Through 6th Kyu*, which gives an insight into both the mental and physical aspects of aikido.

Finally, he's gone out of his way to post extensive supplementary material for this book at www.learn-economics.com. Check it out.

Dedication

To all the family, friends, and colleagues who did so much to set me on the true path, and especially to Andrew Scott, who showed an inordinate amount of faith in me, and Vinetta Archer-Dyer who tirelessly cleared up the mess I left in my wake.

– Peter Antonioni

To my dad, Thomas Ray Flynn, who always impressed upon me the importance of good economic policy both for improving our quality of life and as our last, best hope for lifting billions out of poverty and disease.

– Sean Masaki Flynn

Authors' Acknowledgments

Many people have left their imprints on me as I've traveled through life, and I feel privileged to have benefited from their wisdom and patience along the way.

My parents who impressed on me the return on investment in education.

Paul, who never let me get away with sloppy thinking. Thanks Bro!

Tanya, who taught me how to practise economics as a true art whilst staying sane.

Tim Hames, for three years of the most inspiring tutorials ever.

All the crew at Birkbeck, and especially Professor Ron Smith who actually made econometrics make sense.

Bryan Finn and David Merrick for showing me how it all actually works.

John Cubbin, in whose debt I will be forever, for actually getting me to produce research. Michael Ball for always believing in me, especially when I didn't.

All my colleagues at UCL, especially Richard Pettinger, who got me into this gig, Irene Brunskill, Linda Hesselman, Jane Walker, and Jane Burns Nurse.

All my students at City University and UCL for making a better economist by making me think about how to actually explain this stuff so it makes sense to someone else, because you never actually know that you know things until you try and tell another what you know.

The entire posse, especially the Eggman, the Tenth Emanation, The Church of Uggy, Cap'n Jules, the Krazza, Noel and Leah, Phil and Sunita, Karen, Alice, Mao, Merv Dawg, and everyone. Big Up!

Big shout out to Andrew, AJ, and Heather for beating me up regularly!

And thanks to Rachael at Wiley, for patiently dealing with my somewhat relaxed attitude to deadlines and my totally blockheaded queries.

Very big thanks indeed to Sean Masaki Flynn for an incredibly well-created US original.

Any omissions are my own fault and if you feel you have been unjustly omitted from the thanks you can demand a pint anytime as recompense. I may even buy you one.

– Peter Antonioni

I'd like to thank the many great economists who managed to get things into my head despite my very thick skull.

Among my teachers, I can't help but thanks Caroline Betts, Tim Cason, Richard Ciccetti, Michael DePrano, Richard Easterlin, Robert Kalaba, Timur Kuran, Jeffrey Nugent, and Morton Shapiro for the excellent education I received as an undergraduate at the University of Southern California.

I was equally blessed at UC Berkeley, where I got to complete a doctorate under the tutelage of some true intellectual giants including George Akerlof, David Card, J. Bradford DeLong, Jan deVries, Barry Eichengreen, Richard Gilbert, Daniel McFadden, Maurey Obstfeld, Matthew Rabin, David Romer, Christina Romer, and Janet Yellen. It was especially fun when Professors McFadden and Akerlof won their respective Nobel Prizes during my last two years at Cal.

However, my fellow economics students often did more than my professors to explain things to me when I wasn't getting them. And they continue to educate me even now. So a very heartfelt thank you to Corinne Alexander, Lorenzo Blanco, Mark Carlson, Carlos Dobkin, Tim Doede, Mike Enriquez, Fabio Ghironi, Petra Geraats, Aaron Green, Galina Hale, Alan Marco, Carolina Marquez, Marcelo Moreira, Petra Moser, Marc Muendler, Stefan Palmqvist, Doug Park, Raj Patel, Steve Puller, Desiree Schaan, Doug Schwalm, Mark Stehr, Sam Thompson, Carla Tully, Jeff Weinstein, and Marta Wosinska.

I've also got to thank my students here at Vassar College. You're bright, diligent, and incredibly hardworking. By asking me so many challenging, insightful questions, you've made me a far better economist.

A big thank you to my literary agent Linda Roghaar and my old friend Mike Jones for getting me this book deal. They heard *Dummies* and immediately thought of me.

The entire production team at Wiley also deserve huge praise. All their edits, suggestions, and formatting have turned out a book that's far better than anything I could have come up with on my own.

I also have to deeply thank Dr Robert Harris, whose comments and suggestions have made the text far better than it would have been otherwise.

Finally, I must thank Melissa Lape. She read my copy and made numerous suggestions that helped make *Economics For Dummies* both clear and concise.

If you've had the patience to read this far, you'll also likely have an inclination to go check out www.learn-economics.com, where I've posted lots of supplementary material to accompany *Economics For Dummies*. You just can't get enough, can you?

– Sean Masaki Flynn

Publisher's Acknowledgments

We're proud of this book; please send us your comments through our Dummies online registration form located at www.dummies.com/register/.

Some of the people who helped bring this book to market include the following:

Acquisitions, Editorial, and Media Development

Project Editor: Rachael Chilvers

Commissioning Editor: David Palmer

Assistant Editor: Ben Kemble

Proofreader: Kelly Cattermole

Production Manager: Daniel Mersey

Cover Photos: © K-PHOTOS / Alamy

Cartoons: Ed McLachlan

Composition Services

Senior Project Coordinator: Lynsey Stanford

Layout and Graphics: Joyce Haughey

Proofreaders: John Greenough

Indexer: Sharon Shock

Publishing and Editorial for Consumer Dummies

Diane Graves Steele, Vice President and Publisher, Consumer Dummies

Kristin Ferguson-Wagstaffe, Product Development Director, Consumer Dummies

Ensley Eikenburg, Associate Publisher, Travel

Kelly Regan, Editorial Director, Travel

Publishing for Technology Dummies

Andy Cummings, Vice President and Publisher, Dummies Technology/General User

Composition Services

Debbie Stailey, Director of Composition Services

Contents at a Glance

<i>Introduction</i>	1
<i>Part I: Economics: The Science of How People Deal with Scarcity</i>	7
Chapter 1: What Does Economics Study? And Why Should You Care?	9
Chapter 2: Cake or Ice Cream? Tracking Consumer Choices	25
Chapter 3: Producing the Right Stuff in the Right Way to Maximise Human Happiness	39
<i>Part II: Macroeconomics: The Science of Economic Growth and Stability</i>	63
Chapter 4: Measuring the Macroeconomy: How Economists Keep Track of <i>Everything</i>	65
Chapter 5: Inflation Frustration: Why More Money Isn't Always a Good Thing	89
Chapter 6: Understanding Why Recessions Happen	111
Chapter 7: Fighting Recessions with Monetary and Fiscal Policy	141
<i>Part III: Microeconomics: The Science of Consumer and Firm Behaviour</i>	169
Chapter 8: Supply and Demand Made Easy	171
Chapter 9: Getting to Know <i>Homo Economicus</i> , the Utility-Maximising Consumer	197
Chapter 10: The Core of Capitalism: The Profit-Maximising Firm	217
Chapter 11: Why Economists Love Free Markets and Competition	243
Chapter 12: Monopolies: How Badly Would You Behave If You Had No Competition?	269
Chapter 13: Oligopoly and Monopolistic Competition: Middle Grounds	291
Chapter 14: Property Rights and Wrongs	315
Chapter 15: Market Failure: Asymmetric Information and Public Goods	329
<i>Part IV: The Part of Tens</i>	345
Chapter 16: Ten (Or so) Famous Economists	347
Chapter 17: Ten Seductive Economic Fallacies	355
Chapter 18: Ten Economic Ideas to Hold Dear	363
<i>Appendix: Glossary</i>	369
<i>Index</i>	377

Table of Contents

.....

<i>Introduction</i>	1
About This Book	1
Conventions Used in This Book	3
What You're Not to Read	3
Foolish Assumptions	4
How This Book Is Organised	4
Part I: Economics: The Science of How People Deal with Scarcity ...	4
Part II: Macroeconomics: The Science of Economic Growth and Stability	5
Part III: Microeconomics: The Science of Consumer and Firm Behaviour	5
Part IV: The Part of Tens	5
Icons Used in This Book	5
Where to Go from Here	6

<i>Part I: Economics: The Science of How People Deal with Scarcity</i>	7
---	----------

Chapter 1: What Does Economics Study? And Why Should You Care? . . . 9

Considering a Little Economic History	10
Pondering just how nasty, brutish and short life used to be	10
Identifying the institutions that led to higher living standards	11
Looking towards the future	12
Sending Macroeconomics and Microeconomics to Separate Corners ...	13
Framing Economics as the Science of Scarcity	13
Zooming Out: Macroeconomics and the Big Picture	14
Measuring the economy	14
Recognising what causes recessions	15
Fighting recessions with monetary and fiscal policies	15
Getting up Close and Personal: Microeconomics	16
Balancing supply and demand	16
Considering why competition is so great	17
Examining problems caused by lack of competition	17
Reforming property rights	18
Dealing with other common market failures	18
Understanding How Economists Use Models and Graphs	19
Abstracting from reality is a good thing	19
Introducing your first model: The demand curve	19
Drawing your own demand curve	22

Chapter 2: Cake or Ice Cream? Tracking Consumer Choices 25

Considering a Model of Human Behaviour.....	26
Maximising Happiness Is the Objective.....	27
Using utility to measure happiness.....	27
Taking altruism and generosity into account.....	28
Realising that self-interest can promote the common good.....	28
Red Light: Examining Your Limitations.....	29
Resource constraints.....	29
Technology constraints.....	30
Time constraints.....	30
Opportunity cost: The unavoidable constraint.....	31
Making Your Final Choice.....	32
Exploring Limitations and Violations of the Economist's Choice Model.....	34
Understanding uninformed decision-making.....	34
Getting rational about irrationality.....	35

Chapter 3: Producing the Right Stuff in the Right Way to Maximise Human Happiness. 39

Reaching the Limit: Determining What's Possible to Produce.....	40
Classifying resources used in production.....	41
Getting less of a good thing: Diminishing returns.....	42
Allocating resources: A little here, a little there.....	43
Graphing your production possibilities.....	44
Pushing the line with better technology.....	47
Determining What Should Be Produced.....	49
Weighing pros and cons of markets and government interventions.....	50
Opting for a mixed economy.....	56
Encouraging Technology and Innovation.....	59

Part II: Macroeconomics: The Science of Economic Growth and Stability..... 63**Chapter 4: Measuring the Macroeconomy: How Economists Keep Track of Everything 65**

Using GDP to Track the Economy.....	66
Leaving some things out of GDP.....	67
Getting in the flow: Tallying up what counts in GDP.....	67
Considering flows of income and assets.....	68
Following the funds, around and around.....	70
Counting stuff when it's made, not when it's sold.....	72
The good, the bad and the ugly: All things increase GDP.....	73

Introducing the GDP Equation	74
C is for consumption	75
I is for investment in capital stock.....	77
G whizz! Government, that is.....	78
NX: Exports less imports	80
Understanding How International Trade Affects the Economy.....	81
Trade deficits can be good for you!.....	81
Considering assets – not just cash	83
Wielding a comparative advantage	85

Chapter 5: Inflation Frustration: Why More Money Isn't Always a Good Thing 89

Buying an Inflation: The Risks of Too Much Money.....	90
Balancing money supply and demand	90
Giving in to the inflation temptation	92
Tallying up the effects of inflation	98
Measuring Inflation: Price Indexes	100
Creating your very own market basket.....	101
Calculating the inflation rate	102
Setting up a price index	103
Determining the real standard of living with the price index	104
Identifying price index problems.....	105
Pricing the Future: Nominal and Real Interest Rates	106
Using the Fisher equation	107
Realising that predictions aren't perfect	107

Chapter 6: Understanding Why Recessions Happen 111

Examining the Business Cycle	112
Striving for Full-Employment Output	113
Returning to Y^* : The Natural Result of Price Adjustments	114
Responding to Economic Shocks: Short-Run and Long-Run Effects	115
Defining some critical terms.....	116
The Tao of P: Looking at price adjustments in the long run	118
A shock to the system: Adjusting to a shift in aggregate demand	119
Dealing with fixed prices in the short run	120
Putting together the long and short of it	123
Heading toward Recession: Getting Stuck with Sticky Prices.....	124
Cutting wages or cutting workers.....	125
Adding up the costs of wages and profits	126
Returning to Y^* with and without government intervention.....	126
Achieving Equilibrium with Sticky Prices: The Keynesian Model.....	127
Adjusting inventories instead of prices	129
Boosting GDP in the Keynesian model.....	138

Chapter 7: Fighting Recessions with Monetary and Fiscal Policy . . . 141

Stimulating Demand to End Recessions	142
Aiming for full-employment output	142
Shifting the AD curve to the right – or, putting people back to work	144
Generating Inflation: The Risk of Too Much Stimulation.....	144
An exercise in futility: Trying to increase output beyond Y^*	145
A temporary high: Tracing the movement of real wages	146
Failing to stimulate: What happens when a stimulus is expected	149
Figuring Out Fiscal Policy	151
Increasing government spending to help end recessions	152
Dealing with deficits	153
Dissecting Monetary Policy.....	156
Identifying the benefits of fiat money.....	156
Realising that you can have too much money!	158
Getting to know bond . . . government bond.....	160
Seeing the link between bond prices and interest rates.....	161
Changing the money supply to change interest rates	162
Lowering interest rates to stimulate the economy	163
Understanding how rational expectations can limit monetary policy	164

**Part III: Microeconomics: The Science of
Consumer and Firm Behaviour 169****Chapter 8: Supply and Demand Made Easy.....171**

Making Sense of Markets	172
Deconstructing Demand	172
Getting our terms straight	173
Graphing the demand curve.....	174
Opportunity costs: Determining the slope of the demand curve	176
Defining demand elasticity	178
Sorting Out Supply	180
Graphing the supply curve	180
Understanding extreme supply cases.....	184
Interacting Supply and Demand to Find Market Equilibrium	185
Finding market equilibrium	185
Demonstrating the stability of the market equilibrium	187
Adjusting to New Market Equilibriums When Supply or Demand Changes	189
Reacting to an increase in demand	189
Reacting to a decrease in supply.....	191
Constructing Impediments to Market Equilibrium	192
Raising price ceilings.....	192
Propping up price floors	193

Chapter 9: Getting to Know *Homo Economicus*, the Utility-Maximising Consumer 197

Knowing the Name of the Game: Constrained Optimisation	198
Finding a Common Denominator to Measure Happiness: Utility	198
Getting Less from More: Diminishing Marginal Utility.....	199
Choosing among Many Options When Facing a Limited Budget.....	202
Trying to buy as much (marginal) utility as you can	203
Allocating money between two goods to maximise total utility	205
Equalising the marginal utility per pound of all goods and services	208
Deriving Demand Curves from Diminishing Marginal Utility	211
Seeing how price changes affect quantities demanded.....	211
Graphing the price and quantity changes to form a demand curve.....	213

Chapter 10: The Core of Capitalism: The Profit-Maximising Firm . . . 217

Maximising Profits Is a Firm's Goal	218
Facing Competition	219
Listing the requirements for perfect competition.....	219
Acting as price takers but quantity makers	220
Distinguishing between accounting profits and economic profits.....	222
Analysing a Firm's Cost Structure	223
Focusing on costs per unit of output	224
Examining average variable costs	226
Watching average fixed costs fall	227
Tracking the movement of average total costs.....	228
Focusing on marginal costs.....	229
Noticing where the MC curve crosses the AVC and ATC curves...	229
Comparing Marginal Revenues with Marginal Costs	231
The magic formula: Finding where $MR = MC$	232
Visualising profits	234
Visualising losses.....	237
Pulling the Plug: When Producing Nothing Is Your Best Bet	238
The short-run shutdown condition: Variable costs exceed total revenues.....	238
The long-run shutdown condition: Total costs exceed total revenues.....	240
At the mercy of the market price.....	241

Chapter 11: Why Economists Love Free Markets and Competition. . . 243

The Beauty of Competitive Free Markets:	
Ensuring That Benefits Exceed Costs.....	244
Examining prerequisites for properly functioning markets	244
Analysing the efficiency of free markets	246
Using total surplus to measure gains	249

When Free Markets Lose Their Freedom:	
Dealing with Deadweight Losses.....	255
Dissecting the deadweight loss from a price ceiling.....	255
Analysing the deadweight loss of a tax.....	256
Hallmarks of Perfect Competition: Zero Profits	
and Lowest Possible Costs	260
Understanding the causes and consequences	
of perfect competition.....	260
Peering into the process of perfect competition.....	261
Graphing how profits guide firm entry and exit.....	263

Chapter 12: Monopolies: How Badly Would You Behave If You Had No Competition?269

Examining Profit-Maximising Monopolies	270
Zeroing in on the problems that monopolies cause	270
Identifying the source of the problem:	
Decreasing marginal revenues	271
Choosing an output level to maximise profits	276
Comparing Monopolies with Competitive Firms.....	280
Looking at output and price levels.....	280
Deadweight losses: Quantifying the harm caused by monopolies ..	282
Focusing on efficiency.....	283
Considering Examples of Good Monopolies	283
Encouraging innovation and investment with patents	284
Reducing annoyingly redundant competitors	284
Keeping costs low with natural monopolies	285
Regulating Monopolies	285
Subsidising a monopoly to increase output.....	286
Imposing minimum output requirements.....	286
Regulating monopoly pricing	287
Breaking up a monopoly into several competing firms	289

Chapter 13: Oligopoly and Monopolistic Competition: Middle Grounds291

Choosing to Compete or Collude	292
Realising that oligopoly firms interact strategically	292
Comparing the outcomes of competition and collusion	293
Cartel Behaviour: Trying to Imitate Monopolists.....	294
Co-ordinating a cartel is hard work.....	294
Examining OPEC to see the difficulties of collusion.....	295
Understanding the Prisoner's Dilemma Model.....	296
Fleshing out the Prisoner's Dilemma	297
Using omerta to resolve the Prisoner's Dilemma	300
Applying the Prisoner's Dilemma to Cartels.....	302
Seeing that OPEC is trapped in a Prisoner's Dilemma	304
Using an enforcer to help OPEC members stick to quotas	305

Regulating Oligopolies	306
Dealing with dominant firms	306
Applying antitrust laws	306
Studying a Hybrid: Monopolistic Competition	307
Benefiting from product differentiation	307
Facing profit limits	308
Chapter 14: Property Rights and Wrongs	315
Allowing Markets to Reach Socially Optimal Outcomes	316
Examining Externalities: The Costs and Benefits	
Others Feel from Our Actions	317
Defining positive and negative externalities	318
Noting the effects of negative externalities	318
Realising that you want positive amounts	
of negative externalities	320
Dealing with negative externalities.....	321
Calculating the consequences of positive externalities	322
Taking in the Tragedy of the Commons	324
Having a cow: Overgrazing on a commonly owned field	324
Sleeping with the fishes: Extinctions	
caused by poor property rights	326
Chapter 15: Market Failure: Asymmetric Information	
and Public Goods	329
Facing Up to Asymmetric Information	330
Realising that asymmetric information limits trade.....	330
Souring on the lemons problem: The used car	
market and the credit crunch.....	331
Issuing insurance when you can't tell individuals apart	335
Providing Public Goods	339
Taxing to provide public goods	340
Enlisting philanthropy to provide public goods.....	341
Providing a public good by selling a related private good.....	341
Ranking new technology as a public good	342
Part IV: The Part of Tens.....	345
Chapter 16: Ten (Or so) Famous Economists	347
Adam Smith	347
David Ricardo.....	348
Karl Marx	348
Alfred Marshall	349
John Maynard Keynes.....	350
Kenneth Arrow and Gerard Debreu	350

Milton Friedman.....	351
Paul Samuelson.....	351
Robert Solow.....	352
Gary Becker.....	352
Robert Lucas.....	353
Chapter 17: Ten Seductive Economic Fallacies	355
The Lump of Labour Fallacy.....	355
The World Is Facing an Overpopulation Problem.....	356
The Fallacy of Confusing Sequence with Causation.....	357
Protectionism Is the Best Solution to Foreign Competition.....	357
The Fallacy of Composition.....	358
If It's Worth Doing, Do It 100 Per Cent.....	359
Free Markets Are Dangerously Unstable.....	360
Low Foreign Wages Mean that Rich Countries Can't Compete.....	360
Tax Rates Don't Affect Work Effort.....	361
Forgetting that Policies Have Unintended Consequences Too.....	362
Chapter 18: Ten Economic Ideas to Hold Dear	363
Society Is Better Off When People Pursue Their Own Interests.....	363
Free Markets Require Regulation.....	364
Economic Growth Depends on Innovation.....	364
Freedom and Democracy Make Us Richer.....	364
Education Raises Living Standards.....	364
Protecting Intellectual Property Rights Promotes Innovation.....	365
Weak Property Rights Cause Many Environmental Problems.....	365
International Trade Is a Good Thing.....	366
Free Enterprise Has a Hard Time Providing Public Goods.....	366
Preventing Inflation Is Easy(ish).....	367
Appendix: Glossary.....	369
Index.....	377

Introduction

Economics is all about humanity's struggle to achieve happiness in a world full of constraints. Too little time and money is available to do everything people want. And things like curing cancer are still impossible because the necessary technologies haven't yet been developed.

But people are clever. They tinker and invent, ponder and innovate. They look at what they have and what they can do with it and take steps to make sure that if they can't have everything, they at least have as much as possible.

Making trade-offs is key. Because you can't have everything, you have to make choices. For example, you have to choose whether to save or spend, whether to stay in school or get a job, and whether the government should spend more money on primary education or on cancer research.

Choice is a fundamental part of everyday life. The science that studies *how* people choose – economics – is indispensable if you really want to understand human beings both as individuals and as members of larger organisations.

Sadly, though, economics has typically been explained so badly that people dismiss it as impenetrable gobbledygook or stand falsely in awe of it – after all, if economics is hard to understand, it must be important, right?

We wrote this book so that you can quickly and easily understand economics for what it is – a serious science that studies a serious subject and has developed some seriously effective ways of explaining human behaviour out in the (very serious) real world. Read this book to understand more about people, government, international relations, business, and even environmental issues such as global warming and endangered species. Economics touches on nearly everything, so the returns on reading this book are huge.

About This Book

Reading this book enables you to discover the most important economic theories, hypotheses, and discoveries without a zillion obscure details, outdated

examples, or complicated mathematical ‘proofs’. Among the topics covered are the following:

- ✔ How the government fights recessions and unemployment using monetary and fiscal policy.
- ✔ How and why international trade is good for us.
- ✔ Why poorly designed property rights are responsible for environmental problems such as global warming, pollution, and species extinctions.
- ✔ How profits guide businesses to produce the goods and services we take for granted.
- ✔ Why competitive firms are almost always better for society than monopolies.
- ✔ How the Bank of England controls the money supply, interest rates and inflation all at the same time.
- ✔ Why government policies such as price controls and subsidies typically cause much more harm than good.
- ✔ How the simple supply and demand model can explain the prices of everything from comic books to open-heart surgeries.

We do our best to explain these things, and much more, clearly and directly. We also structure this book to put *you* in control. You can read the chapters in any order, and you can immediately jump to what you need to know without having to read a bunch of stuff that you couldn’t care less about.

Economists like competition, so don’t be surprised that a lot of competing views and paradigms exist among economists. Indeed, only through vigorous debate and careful review of the evidence can the profession improve its understanding of how the world works.

In this book, we try to steer clear of fads or ideas that foster a lot of disagreement. This book contains core ideas and concepts that economists agree are useful and important. (If you want to be subjected to our personal opinions and pet theories, you’re going to have to buy us a drink.)

However, economists have honest disagreements about how to present even the core concepts, so we had to make some decisions about organisation and structure. For example, we present macroeconomics using a Keynesian framework even when we explain some rather non-Keynesian concepts. (You don’t need to worry if you don’t know who this Keynes fellow is or what makes him so *Keynesian*, because we introduce him to you later in the book.) Some people may quibble with this approach, but we think it makes for a succinct presentation.

Conventions Used in This Book

Economics is full of two things you may not find very appealing: jargon and algebra. To minimise confusion, whenever we introduce a new term, we put it in *italics* and follow it closely with an easy-to-understand definition. Also, whenever we bring algebra into the discussion, we use those handy *italics* again to let you know that we're referring to an algebraic element. For example, *I* indicates investment, so you may see a sentence like this one: You may say to yourself, 'I think that *I* is too big.'

We try to keep equations to a minimum, but sometimes they actually help to make things clearer. In such instances, we sometimes have to use several equations one after another. To avoid confusion about which equation we refer to at any given time, we give each equation a number, which we put in parentheses. For example,

$$\text{happy} = \text{beer} + \text{curry}^2 \quad (1)$$

Finally, the following conventions are used throughout the text of all *For Dummies* books to make things consistent and easy to understand:

- ✓ All web addresses appear in `this font`.
- ✓ **Bold** is used to highlight the action parts of numbered steps.

What You're Not to Read

The whole point of a *For Dummies* book is to give you quick access to the essentials so that you don't have to wade through a bunch of stories, facts, and anecdotes. On the other hand, sometimes stories, facts, and anecdotes can be both fun and enlightening.

But even when they are fascinating, doesn't mean you should be forced to read them. Consequently, we clearly identify all the 'skippable' material. This information is the stuff that, although interesting and related to the topic at hand, isn't essential for you to know:

- ✓ **Text in sidebars:** The sidebars are shaded boxes that share interesting stories and observations, but aren't necessary reading.
- ✓ **The bits on the acknowledgements page:** Unless you're one of our friends who needs an ego boost, nothing's here for you.

Naturally, we'd like to believe that you're going to choose to read *everything* we've written, but don't worry: we aren't going to find out.

Foolish Assumptions

We wrote this book assuming some things about you:

- ✔ You're sharp, thoughtful, and interested in how the world works.
- ✔ You're a secondary school or university student trying to flesh out what you're learning in class, or you're a citizen of the world who realises that a good grounding in economics is going to help you understand everything from business and politics to social issues such as poverty and environmental degradation.
- ✔ You want to know some economics, but you're also busy leading a very full life. Consequently, although you want the crucial facts, you don't want to have to read through a bunch of minutia to find them.
- ✔ You're not totally intimidated by numbers, facts, and figures. Indeed, you welcome them because you like to have things proven to you rather than taking them on faith because some pinhead with a PhD says so.
- ✔ You like discovering *why* as well as *what*. That is, you want to know why things happen and how they work rather than just memorising facts.
- ✔ Finally, you're better-looking than average and have a good sense of style. In particular, you really love this book's stylish yellow and black cover and feel almost hypnotically compelled to buy a copy.

How This Book Is Organised

This book is divided into four parts to make the material easier to understand and access. Part I covers the big concepts that motivate how economists look at the world. Parts II and III follow the traditional division of economics into two halves: *Macroeconomics* deals with big-picture issues like recessions and international trade, whereas *microeconomics* focuses on individual people, businesses, and industries. Part IV is The Part of Tens and contains a few fun but informative top-ten lists.

Part I: Economics: The Science of How People Deal with Scarcity

Economics is all about how people deal with scarcity. Too little time is available, and only a finite supply of natural resources such as oil and iron. Consequently, people have to be clever about getting the most out of life – choosing wisely about what to do with the limited resources they're given. Part I explains how people go about dealing with scarcity and the trade-offs

that it forces them to make. The rest of economics is just seeing how scarcity forces people to make trade-offs in more specific situations.

Part II: Macroeconomics: The Science of Economic Growth and Stability

Macroeconomics views the economy from on high, at the national or international level, and deals with the choices that countries face about economic growth and development and about how to best manage their economies to avoid recessions. Macroeconomics also deals with the misery caused by things such as unemployment and inflation. In this part, you find out about monetary and fiscal policy, the Bank of England, the effects of taxation on the economy, and international trade and trade policy.

Part III: Microeconomics: The Science of Consumer and Firm Behaviour

Microeconomics focuses on the behaviour of individual people and individual firms, studying what motivates them and how they act to achieve their goals given the constraints they face. In this part, you discover what motivates firms to produce output, how buyers and sellers interact in markets to allocate that output, and how markets can break down and do perverse things if not properly managed. You also find out about supply and demand, competition, monopolies, Adam Smith's invisible hand, and lots of nifty applications of economics to things such as insurance markets and environmental issues. Economics really does get into everything.

Part IV: The Part of Tens

Every *For Dummies* book ends with top-ten lists that are both helpful and fun. In this part, we give you short bios of famous economists (explaining what they discovered and why it was so important), economic ideas to hold dear, and false economic assertions that you probably hear repeated all the time in the media and by self-serving politicians.

Icons Used in This Book

To make this book easier to read and simpler to use, we include a few icons that can help you find and fathom key ideas and information.



Theories are always easier to understand with an example. So when you see this icon you know you're in for some help, usually using everyday items like pizza and beer. (We find pizza and beer help in all sorts of ways.)



This icon alerts you that we're explaining a really fundamental economic concept or fact. It saves you the time and effort of marking up the book with a highlighter.



Sometimes we find it helpful to kick theories out into the real world to see how they actually work. This icon alerts you that a helpful real-world application is nearby.

Where to Go from Here

This book is set up so that you can jump in anywhere and understand what you're reading. For example:

- ✔ Want the knowledge on how the Bank of England changes interest rates to stimulate the economy and fight recessions? Jump right to Chapter 7.
- ✔ Want to know about environmental economics and how most environmental problems are caused by poorly designed property rights? Open the book at Chapter 14.
- ✔ Need to figure out why everyone talks about supply and demand? Hit Chapter 8.

The book is also divided into independent parts so that you can, for example, read all about microeconomics without having to read anything about macroeconomics. And the table of contents and index can help you find specific topics easily.

But, hey, if you don't know where to begin, just do the old-fashioned thing and start at the beginning. As that popular song from the film *The Sound of Music* says, 'Let's start at the very beginning! A very good place to start.'

Part I

Economics: The Science of How People Deal with Scarcity



'It used to be called "The Economic
Miracle" in the boom days.'

In this part . . .

Economics studies how people deal with scarcity and the inescapable fact that our wants typically exceed the means available to satisfy them. The fact that life has limits may not at first seem like a good basis for an entire social science, but every government decision, every business decision and a large chunk of your personal decisions all come down to deciding how to get the most out of limited resources. Consequently, as we explain in this part, economics is fundamental to almost all aspects of life.

Chapter 1

What Does Economics Study? And Why Should You Care?

In This Chapter

- ▶ Taking a quick peek at economic history
 - ▶ Observing how people cope with scarcity
 - ▶ Separating macroeconomics and microeconomics
 - ▶ Growing the economy and avoiding recessions
 - ▶ Understanding individual and firm behaviour
 - ▶ Getting a grip on the graphs and models that economists love to use
-

Economics is the science that studies how people and societies make decisions that allow them to get the most out of their limited resources. Because every country, every business and every person deals with constraints and limitations, economics is literally everywhere.

For example, you could be doing something else right now instead of reading this book: exercising, watching a film or talking with a friend. The only reason for you to be reading this book is that doing so is the best possible use of your very limited time.

In the same way, you hope that the paper and ink used to make this book have been put to their very best use. Similarly, we all hope that every last tax pound that the government spends is being used in the best possible way, and not being dissipated on projects of secondary importance.

Economics gets to the heart of these issues, analysing individual and firm behaviour, as well as social and political institutions, to see how well they perform at converting humanity's limited resources into the goods and services that best satisfy human wants and needs.

Considering a Little Economic History

To better understand today's economic situation and what sort of policy and institutional changes may promote the greatest improvements, you have to look back on economic history to see how humanity arrived at its current situation. Stick with us: we make this brief survey as painless as possible for any history haters.

Pondering just how nasty, brutish and short life used to be

For most of human history, people didn't manage to squeeze much out of their limited resources. Standards of living were quite low, and people lived poor, short and rather painful lives. Consider the following facts, which didn't change until just a few centuries ago:

- ✓ Life expectancy was about 25 years.
- ✓ More than 30 per cent of newborns never reached their fifth birthdays.
- ✓ Women had a 10 per cent chance of dying during childbirth.
- ✓ Most people had personal experience of horrible diseases and/or starvation.
- ✓ The standard of living for one generation was no higher than that of previous generations. Except for the nobility, everybody lived at or near subsistence level, century after century.

In the last 250 years or so, however, everything changed. A process of rapid innovation led to the invention or exploitation of electricity, engines, complicated machines, computers, radio, television, biotechnology, scientific agriculture, antibiotics, aviation and a host of other technologies. Each of these items enabled humankind to do much more with the limited amounts of air, water, soil and sea available on planet earth.

The result was an explosion in living standards, with life expectancy at birth now well over 60 years worldwide and many people able to afford much better housing, clothing and food than was even imaginable a few hundred years ago.

Of course, not everything is perfect. Grinding poverty is still a fact of life in a large portion of the world, and even the richest nations have to cope with pressing economic problems like unemployment, persistent poverty or lack of access to resources.



But the fact remains that the modern world is a much richer place than it has been in the past, and we now have sustained economic growth in most nations, which means that living standards have been rising consistently year after year.

Identifying the institutions that led to higher living standards

The obvious reason for higher living standards, which continue to rise, is that human beings have recently figured out lots of new technologies, and we keep inventing more. But if you dig a little deeper, you have to wonder why a technologically innovative society didn't happen earlier.

The ancient Greeks invented a simple steam engine and the coin-operated vending machine. They even developed the basic idea behind the programmable computer. But they never quite got around to having an industrial revolution and embarking on a path of sustained economic growth.

And despite the fact that every society has had its share of really smart people, it wasn't until the late 18th century, in England, that the Industrial Revolution got started and living standards in many nations rose substantially and kept on rising, year after year.

So what factors combined in the late 18th century to accelerate economic growth so radically? The short answer is that the following institutions were in place:

- ✔ **Democracy:** Yes, we can overstate this one. However, it does make sense to consider that you're more likely to invest if your investment is protected by the rule of law instead of depending on the whim of a tyrant. Also, the governments of the democratic era have been better at incorporating the views of the merchants and manufacturers who have created the wealth that we now enjoy.
- ✔ **The limited liability corporation:** Under this business structure, investors would lose only the amount of their investment and not be liable for any debts that the corporation was unable to pay. Limited liability greatly reduced the risks of investing in businesses and, consequently, led to much more investing.

- ✔ **Patent rights to protect inventors:** Before patents, inventors usually saw their ideas stolen before they made any money. By giving inventors the exclusive right to market and sell their inventions, patents gave a financial incentive to produce lots of inventions. Indeed, after patents came into existence, the world saw its first full-time inventors – people who made a living inventing things.
- ✔ **Widespread literacy and education:** Without highly educated inventors, new technologies don't get invented. And without an educated workforce, the invented products can't be mass-produced. Consequently, the decision that many nations made to make primary and then secondary education mandatory paved the way for rapid and sustained economic growth.

Institutions and policies like these gave us a world of growth and opportunity, an abundance so unprecedented in world history that the greatest public health problem in many countries today is obesity.

Looking towards the future

The world faces many challenges now, some of which are the results of our successes and some are about extending them to all the world's citizens. Among the former are the potential changes in climate and our reaction to them, and among the latter the question of how we improve the condition of poorer citizens without creating environmental damage.

Some problems, like grinding poverty, can be alleviated by extending to poorer nations the institutions that have already been proven in richer nations to lead to rising living standards. But other problems, like the pollution and resource depletion that come with the institutional structures used in richer nations, require new inventions and new institutions.

Consequently, we offer two related and very good reasons for you to read this book and find out about economics:

- ✔ You'll discover how modern economies function, providing you with an understanding not only of how they've so greatly raised living standards, but also of where they need some improvement.
- ✔ You'll get a thorough grasp of fundamental economic principles, allowing you to judge for yourself the economic policy proposals that politicians and others promote. After reading this book, you'll be able to sort the good from the bad.

Sending Macroeconomics and Microeconomics to Separate Corners

We organise this book to get as much economics into you as quickly and effortlessly as possible. We've also done our best to keep the subject lively and fun. The Scottish historian and essayist Thomas Carlyle called economics the 'dismal science' (he was a bit of a wag), but that's not really fair, and we're going to do our best to make sure that you don't come to agree with him.



The main organising principle we use in this book is to divide economics into two broad pieces, macroeconomics and microeconomics:

- ✓ *Macroeconomics* looks at the economy as an organic whole, concentrating on economy-wide factors such as interest rates, inflation and unemployment. Macroeconomics also encompasses the study of economic growth and how governments use monetary and fiscal policy to try to moderate the harm caused by recessions.
- ✓ *Microeconomics* focuses on individual people and individual businesses. Microeconomics explains how individuals behave when faced with decisions about where to spend their money or how to invest their savings, and how profit-maximising firms behave both individually and when they're competing against each other in markets.

Underlying both macroeconomics and microeconomics are some basic principles such as scarcity and diminishing returns. Consequently, we spend the rest of Part I explaining these fundamentals before diving into macroeconomics in Part II and microeconomics in Part III.

Most of the rest of this chapter serves as a trailer for the rest of the book, so if you want to be surprised later on, flip ahead a few pages right now. The exception is the last section, where we talk about how economists use charts and graphs. If you need to brush up on how to read charts and graphs, read that section before jumping into other chapters.

Framing Economics as the Science of Scarcity

Scarcity is the fundamental and unavoidable phenomenon that creates a need for the science of economics. Without scarcity of time, scarcity of resources,

scarcity of information, scarcity of consumable goods and scarcity of peace and goodwill on earth, human beings would lack for nothing.

Scarcity is why you can't have everything, even if you're the richest person in the world. Even if money's not scarce, time and/or physical resources will be. At some level, you're going to make choices about what you spend all that lovely lolly on. An economist is keen to point out that you just can't have it all!



Sadly, scarcity is a fact. Too little time and stuff exists to satisfy all our desires, so people have to make hard choices about what to produce and consume; if they can't have everything, they at least have the best that's possible under the circumstances. Chapter 2 gets deep into scarcity and the trade-offs that it causes people to make.

Chapter 3 builds on Chapter 2 by showing you how economists analyse the decisions that people make about how to best maximise human happiness in a world of scarcity. That process turns out to be intimately connected with a phenomenon known as *diminishing returns*, which describes the sad fact that each additional amount of a resource that's thrown at a production process brings forth successively smaller amounts of output.

Like scarcity, diminishing returns is unavoidable, and in Chapter 3 we explain how people very cleverly deal with this phenomenon in order to get the most out of humanity's limited pool of resources.

Zooming Out: Macroeconomics and the Big Picture

Part II of this book covers macroeconomics, which treats the economy as a unified whole. Studying macroeconomics is useful because certain factors, such as interest rates and tax policy, have economy-wide effects, and also because when the economy goes into a recession or a boom, every person and every business is affected. Macroeconomics is the stuff of the big picture that gets reported on the news.

Measuring the economy

In Chapter 4, we show you how economists measure *gross domestic product (GDP)*, the value of all goods and services produced in the economy in a given period of time, usually a quarter or a year. Totalling up this number is

absolutely vital because if you can't measure how the economy is doing, you can't tell whether government policies intended to improve the economy are helping or hurting.

Inflation measures how prices in the economy change over time. This topic, which is the focus of Chapter 5, is crucial because high rates of inflation usually accompany huge economic problems, including deep recessions and countries defaulting on their debts.

Studying inflation is also important because poor government policy is the sole culprit behind high rates of inflation – meaning that governments are totally responsible when large inflation rates happen.

Recognising what causes recessions



Recessions linger only because institutional factors in the economy make it very hard for prices in the economy to fall. As we explain in Chapter 6, if prices *could* fall quickly and easily, recessions would quickly resolve themselves. But because prices can't quickly and easily fall, economists have had to develop anti-recessionary policies to help get economies out of recessions as quickly as possible.

Fighting recessions with monetary and fiscal policies

The man most responsible for developing anti-recessionary policies was the English economist John Maynard Keynes, who in 1936 wrote the first macroeconomics book about fighting recessions. Chapter 6 introduces you to his model of the economy and how it explicitly takes account of the fact that prices can't quickly and easily fall to get you out of recessions. Because his economic model takes that fact into account, it serves as the perfect vehicle for illustrating the two things that *can* help get you out of a recession.

These two things are monetary and fiscal policy, which are covered in-depth in Chapter 7:

- ✓ **Monetary policy** uses changes in the money supply to change interest rates in order to stimulate economic activity. For example, if the government causes interest rates to fall, consumers borrow more money to buy things like houses and cars, thereby stimulating economic activity and helping to get the economy moving faster.

✔ **Fiscal policy** refers to using increased government spending or lower tax rates to help fight recessions. For example, if the government buys more goods and services, economic activity increases. In a similar fashion, if the government cuts tax rates, consumers end up with higher after-tax incomes, which, when spent, increase economic activity.

In the first decades after Keynes's anti-recessionary ideas were put into practice, they seemed to work really well. However, they didn't fare so well during the 1970s, and it became apparent that although monetary and fiscal policy were powerful anti-recessionary tools, they had their limitations.

For this reason, Chapter 7 also covers how and why monetary and fiscal policy are constrained in their effectiveness. The key concept is called *rational expectations*. This concept explains how rational people very often change their behaviour in response to policy changes in ways that limit the effectiveness of those changes. You need to understand this concept if you're going to come up with informed opinions about current macroeconomic policy debates.

Getting up Close and Personal: Microeconomics

Although macroeconomics is concerned with government policies to improve the overall economy, microeconomics gets down to the nitty-gritty, studying the most fundamental economic agents: individuals and firms.

Balancing supply and demand

In a modern economy, individuals and firms produce and consume everything that gets made. Consequently, Part III's coverage of microeconomics begins in Chapter 8 by focusing on how supply and demand determine prices and output levels in competitive markets. Supply and demand is a logical place to begin because producers determine supply, consumers determine demand and their interaction in markets determines what gets made and how much it costs.

Chapter 9 digs in deeper to see how individuals make economic decisions about how to get the most happiness out of their limited incomes. These decisions generate the demand curves that affect prices and output levels in markets.

In a similar way, the profit-maximising decisions of firms generate the supply curves that affect markets. In Chapter 10, we explain how that happens, and we also explain how profit-maximising firms actually go about maximising their profits. If you've ever had some nasty thoughts about capitalism – and who hasn't? – this chapter puts you eyeball-to-eyeball with the enemy.

Considering why competition is so great

You may not feel warm and fuzzy about profit-maximising firms, but economists love them – just as long as the firms are stuck in competitive industries. The reason, briefly, is that firms that are forced to compete end up satisfying two wonderful conditions:

- ✓ Competitive firms are *allocatively efficient*, which simply means that they produce the goods and services that consumers most greatly desire to consume.
- ✓ Competitive firms are *productively efficient*, which means that they produce these goods and services at the lowest possible cost.



These two great facts about competitive firms are the basis of Adam Smith's famous *invisible hand* – the idea that when constrained by competition, each firm's greed ends up causing it to act in a socially optimal way, as if guided to do the right thing by an invisible hand. We discuss this idea, and much more about the benefits of competition, in Chapter 11.

Examining problems caused by lack of competition

Unfortunately, not every firm is constrained by competition. And when that happens, firms don't end up acting in socially optimal ways.

The most extreme case is a *monopoly*, a situation in which only one firm exists in an industry – meaning that it has absolutely no competition. As we explain in Chapter 12, monopolies behave very badly, restricting output in order to drive up prices and inflate profits. These actions, which hurt consumers, go on indefinitely unless a government takes steps to regulate the firm's behaviour.

A less extreme case of lack of competition is *oligopoly*, a situation in which only a few firms exist in an industry. In such situations, firms could make deals not to compete against each other so that they can keep prices high

and make bigger profits. To prevent that, most countries have laws making such price fixing illegal.

In Chapter 13, we examine oligopoly firms in depth. We explain not only how they misbehave, but also the fact that they often have a hard time keeping their agreements with each other to maintain high prices and high profits. This fact means that oligopoly firms often end up competing against each other despite their best efforts not to. Consequently, government regulation isn't always needed.

Reforming property rights

Markets and competition can be relied upon to produce socially beneficial results only if society sets up a good system of property rights. Almost all pollution issues, as well as all cases of loss of animal species, are the direct result of poorly designed property rights generating perverse incentives to do bad things. Economists take this problem very seriously and have done their best to reform property rights in order to alleviate pollution and reduce the threat of animal species extinction. We discuss these issues in detail in Chapter 14.

Dealing with other common market failures

Monopolies, oligopolies and poorly designed property rights all lead to what economists like to call *market failures* – situations where markets deliver socially non-optimal outcomes. Two other common causes of market failure are asymmetric information and public goods:

- ✓ *Asymmetric information* refers to situations in which the buyer knows more than the seller about the quality of the goods being negotiated over, or alternatively the seller knows more than the buyer. Because of the uneven playing field between the two parties and the suspicions it creates, a lot of potentially beneficial economic transactions never get completed.
- ✓ *Public goods* refer to goods or services that are impossible to provide to just one person; if you provide them to one person, you have to provide them to everybody. (Think of a fireworks display, for example.) The problem is that most people try to get the benefit without paying for it.

We discuss both these situations, and ways to deal with them, in Chapter 15.

Understanding How Economists Use Models and Graphs

Economists like to be logical and precise, which is why they use a lot of algebra and maths. But they also like to present their ideas in easy-to-understand and highly intuitive ways, which is why they use so many graphs. To avoid a graph-induced panic as you flip through the pages of this book, we want to spend a few pages helping you get acquainted with what you're going to encounter in other chapters. Take a deep breath; we promise this isn't going to hurt.

Abstracting from reality is a good thing

Economists use graphs that are almost always visual representations of economic models. An *economic model* is a mathematical simplification of reality that allows you to ignore all the irrelevant details in order to focus on what's really important.

For example, the economist's model of consumer demand focuses on how prices affect the amounts of goods and services that people want to buy. Obviously, other things, such as changing styles and tastes, affect consumer demand as well, but price is key. Consider orange juice, for example. The price of orange juice is the major thing that affects how much orange juice people are going to buy. (We don't care what dietary trend is in vogue – if orange juice costs £50 a litre, you're probably going to find another diet.) Therefore, abstracting from those other things is helpful and allows you to concentrate solely on how the price of orange juice affects the quantity of orange juice that people want to buy.

Introducing your first model: The demand curve



Suppose that economists go out and survey consumers, asking them how many litres of orange juice they would buy each month at three hypothetical prices: £10 per litre, £5 per litre and £1 per litre. The results are summarised in Table 1-1.

Table 1-1 Litres of Orange Juice That Consumers Want to Buy

<i>Price</i>	<i>Litres</i>
£10	1
£5	6
£1	10

Economists refer to the quantities that people are willing to purchase at various prices as the *quantity demanded*, or *the demand*, at those prices. When you look at the data in Table 1-1, you find that the price of orange juice and the quantity demanded of orange juice have an *inverse relationship* with each other – meaning that when one goes up, the other goes down.



Because this inverse relationship between price and quantity demanded is so universal and holds true for nearly all goods and services, economists refer to it as the *Law of Demand*. But, quite frankly, the Law of Demand becomes much more immediate and interesting if you can *see* it rather than just think about it.

Creating the demand curve by plotting out data

The best way to *see* the data in Table 1-1 is to plot it out on a chart. In Figure 1-1, we mark three points and label them *A*, *B* and *C*. The horizontal axis of Figure 1-1 measures the number of litres of orange juice that people demand each month at various prices per litre. The vertical axis measures the prices.

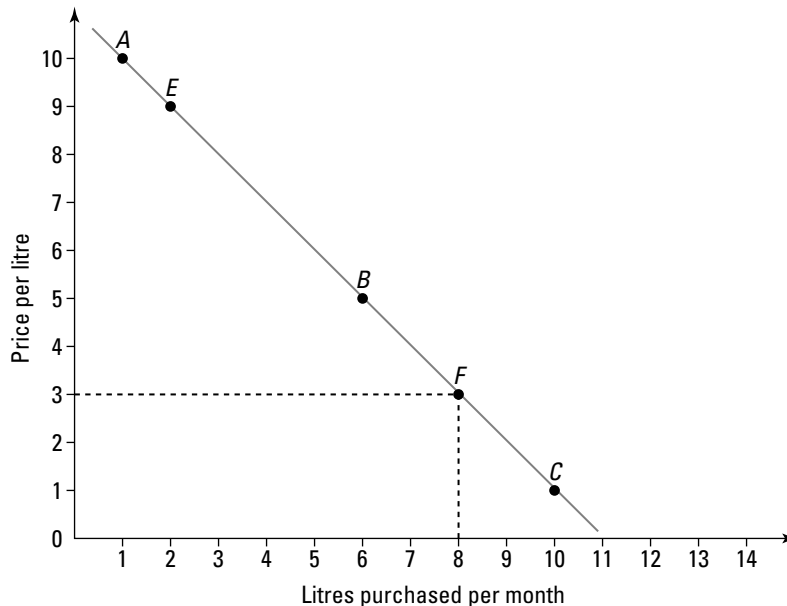


Figure 1-1:
Graphing
the demand
for orange
juice.

Point *A* is the visual representation of the data in the top row of Table 1-1. Point *A* tells you that at a price of £10 per litre, people want to purchase only 1 litre per month of orange juice. Similarly, point *B* tells you that they demand 6 litres per month at a price of £5, while point *C* tells you that they demand 10 litres per month at a price of £1 per litre.

Notice that we connect the points *A*, *B* and *C* with a line. We do this to account for the fact that the economists who conducted the survey asked about what people would do at only three prices. If they had had a big enough budget to ask consumers about every possible price (£8.46 per litre, £2.23 per litre and so on), the graph would have an infinite number of dots. But because they didn't do that, we interpolate by drawing a straight line. The line does a pretty good job of estimating what people's demands are for prices that the economists didn't survey.

The straight line connecting the points in Figure 1-1 is called a *demand curve*. We know it doesn't curve at all, but for simplicity, economists use the term *demand curve* to refer to all plotted relationships between price and quantity demanded, regardless of whether they're straight lines or curvy lines. (This convention is consistent with the fact that economists are both eggheads *and* squares.)

Straight or curvy, you can now *visualise* the fact that price and quantity demanded have an inverse relationship. The inverse relationship implies that demand curves slope downward. You can now see that when price goes up, quantity demanded goes down.

Using the demand curve to make predictions

Graphing out the demand curve also allows for a much greater ability to make quick predictions. For example, the straight line in Figure 1-1 can be used to estimate that at a price of £9 per litre, people are going to want to buy about 2 litres per month of orange juice. We label this point *E* on the graph.

Suppose that you can only see the data in Table 1-1 and can't look at Figure 1-1. Would you be able to estimate quickly how many litres per month people are likely to demand if the price of orange juice is £3 per litre? Looking at the second and third rows of Table 1-1, you have to conclude that people are going to demand somewhere between 6 and 10 litres per month. But working out exactly how many litres are going to be demanded would take some time and require some annoying algebra.

If you look at Figure 1-1, working out how many litres per month people are going to demand is easy. You just start at the price £3 on the vertical axis, move sideways to the right until you hit the demand curve at point *F*, and drop down vertically until you get to the horizontal axis, where you discover that you're at 8 litres per month. (To clarify what we mean, we drew in a dotted line in Figure 1-1 that follows this path.)

As you can see, using a graph rather than a table makes coming up with model-based predictions much, much simpler.

Drawing your own demand curve

To make sure that you're comfortable using graphs, we encourage you to do a simple exercise that involves plotting some points and drawing lines between them. That's not so hard, right?



Imagine that the government came out with a research report showing that people who drink orange juice have lower blood pressure, fewer strokes and a better sex life than people who don't drink orange juice. What do you think is likely to happen to the demand for orange juice? Obviously, it should increase.

To verify this, our intrepid team of survey economists goes out once again and asks people how much orange juice they're going to buy each month at each of the three prices listed in Table 1-1: £10, £5, and £1. The new responses are given in Table 1-2.

Table 1-2 Litres of Orange Juice That Consumers Want to Buy after Reading New Government Research

<i>Price</i>	<i>Litres</i>
£10	4
£5	9
£1	13

Your assignment, should you choose to accept it, is to plot these three points on Figure 1-1. After you've done that, connect them with a straight line. (Yes, you can write in the book!)

What you've just created is a new demand curve that reflects people's new preferences for orange juice in light of the government survey. Their increased demand is reflected in the fact that at any given price, they now demand a larger quantity of juice than they did before. For example, whereas before they wanted only 1 litre per month at a price of £10, they're now willing to buy 4 litres per month at that price.

Of course, an inverse relationship still exists between price and quantity demanded, meaning that even though the health benefits of orange juice make people demand more orange juice, people are still sensitive to higher orange juice prices. Higher prices still mean lower quantities demanded, and your new demand curve still slopes downward.

Ready for one last exercise before you dive into the rest of the book? Use your new demand curve to figure out how many litres per month people are now going to want to buy at a price of £7 and at a price of £2. Working these things out from the data in Table 1-2 would be hard, but figuring them out using your new demand curve should be easy.

Chapter 2

Cake or Ice Cream? Tracking Consumer Choices

In This Chapter

- ▶ Deciding what brings the most happiness
 - ▶ Cataloguing the constraints that limit choice
 - ▶ Modelling choice behaviour like an economist
 - ▶ Evaluating the limitations of the choice model
-

Economics is all about *how* groups and individuals make choices and *why* they choose the things that they do. Economists have spent a great deal of time analysing how groups make choices, but because group choice behaviour usually turns out to be very similar to individual choice behaviour, our focus in this chapter is on individuals.

To keep things simple, our explanation of individual choice behaviour focuses on *consumer behaviour*, because most of the choices people make on a day-to-day basis involve which goods and services to consume. But, of course, real-life choices often encompass a lot of other things, some of them very weighty. For example, people must make choices about long-term things like whether to get a job or continue in education, as well as things of the greatest possible seriousness like whether to continue negotiating or declare war.

Human beings are constantly forced to choose because our wants almost always exceed our means. Limited resources, or *scarcity*, is at the heart not only of economics, but also of ecology and biology. Darwinian evolution is all about animals and plants competing over limited resources to produce the greatest number of progeny. Economics is about human beings choosing among limited options to maximise happiness.

Considering a Model of Human Behaviour

Human beings may be complicated creatures with sometimes mystifying behaviour, but at root, we can understand their choices in a fairly universal way. As a result, a lot can be gained by studying choice behaviour because if we can understand the *how and why* of the choices people made in the past, we stand a very good chance of understanding the choices they're going to make in the future.

Understanding (and even, sometimes, predicting) future choice behaviour is very important because major shifts in the economic environment are typically the result of millions of small individual decisions that add up to a major trend. For example, the circumstances that cause millions of individuals to choose between pursuing work or further education accumulate over the years, producing major effects on the unemployment rate. And the choices these individuals make about how much of their pay cheques to save or spend determine whether interest rates rise or fall, and also whether gross domestic product (GDP) and overall economic output increase or decrease. (We discuss GDP in Chapter 4.)

In order to predict how self-interested individuals make their choices, economists have created a model of human behaviour that assumes rationality and the ability to calculate subtle trade-offs between possible choices. This model is a three-stage process:

- 1. Evaluate how happy each possible option can make you.**
- 2. Look at the constraints and trade-offs limiting your options.**
- 3. Choose the option that maximises your overall happiness.**

Although not a fully complete description of human choice behaviour, this model generally makes accurate predictions. However, many people question this explanation of human behaviour. Here are three common objections:

- ✔ Are people really so self-interested? Aren't people often motivated by what's best for others?
- ✔ Are people really aware at all times of all their options? How are they supposed to choose rationally among new things that they have never tried before?
- ✔ Are people really free to make decisions? Aren't they constrained by legal, moral and social standards?

We spend the next few sections of this chapter expanding on the three-step economic choice model and addressing the objections.

Maximising Happiness Is the Objective

Economists like to think of human beings as free agents, with free wills. To economists, people are fully rational and capable of deciding things on their own. But that begs the question of what motivates people and, in turn, of what sorts of things people choose to do given their free wills.

In a nutshell, economists assume that the basic motivation driving most people most of the time is a *desire to be happy*. This assumption implies that people make choices on the basis of whether or not those choices are going to make them as happy as they can be given their circumstances.

Using utility to measure happiness

If people make choices on the basis of what's going to bring them the most happiness, they need a way of comparing how much happiness each option brings. Along these lines, economists assume that people get a sense of satisfaction or pleasure from the things life offers. Sunsets are nice. Eating ice cream is nice. Friendship is nice. And some people happen to like driving fast.



Economists suppose that you can compare all possible things that you may experience with a common measure of happiness or satisfaction, which they call *utility*. Things you like a lot have high utility, whereas things that you like only a little have low or even negative utility.

The concept of utility is very inclusive. For a hedonist, utility may be the physical pleasure enjoyed from experiencing various things. But for a morally conscientious person, utility may be the sense of moral satisfaction that comes with doing the right thing in a given situation. The important thing for economists is that people can ascertain and compare the utilities of various possible activities. Utility acts as a common denominator that allows people to compare even radically different things sensibly.

Taking altruism and generosity into account

Economists take it as a given that people make choices in order to maximise their personal happiness. This viewpoint immediately raises objections because people are often willing to endure great personal suffering in order to help others.

Yet, to an economist, you can view the desire to help others as being a personal preference. The mother who doesn't eat in order to give what little food she has to her infant may be pursuing a goal (helping her child) that maximises the *mother's* own happiness. The same can be said about people who donate to charities. Such generosity, which most people consider to be 'selfless', can also be seen as being consistent with assuming that people do things to make themselves happy. If people donate to charities because doing so makes them feel good, their selfless action is motivated by 'selfish' intention. Because economists see human motivation as selfish, economics is often accused of being immoral. That's not entirely fair.

Economics is concerned with how people achieve their goals instead of questioning the morality of those goals. For example, some people like honey, but others don't. Economists make no distinction between these two groups regarding the rightness or wrongness of their preferences. Instead, what interests economists is how each group behaves given its preferences. Consequently, economics is amoral rather than immoral.

Economists, however, are also people, and they're very concerned with things like social justice, global warming and poverty. They just tend to model the desire to pursue morality and equity as an individual goal that maximises individual happiness rather than as a group goal to be pursued to achieve some sort of collective good.

Realising that self-interest can promote the common good

Adam Smith, one of the fathers of modern economics and the face on the \$20 note, believed that if society is set up correctly, people chasing after their individual happiness provide for other people's happiness as well. As he famously pointed out in *An Inquiry into the Nature and Causes of the Wealth of Nations*, published in 1776, 'It is not from the benevolence of the butcher, the brewer, or the baker, that we can expect our dinner, but from their regard to their own interest.'



To put it bluntly, the butcher, the brewer and the baker don't make stuff for you because they like you, but because they want your money. Yet because they want your money, they end up producing for you everything that you need to have a nice meal. When you trade them your money for their goods, everyone is happier. You think that not having to prepare all that food is worth more to you than keeping your money. And they think that getting your money is worth more to them than the toil involved in preparing all that food.

Adam Smith expanded on this notion by saying that a person pursuing his own selfish interests may be 'led by an invisible hand to promote an end which was no part of his intention'. Because economists recognise this 'invisible hand', they're less concerned with intent than with outcome, and less concerned with what makes people happy than with how they pursue the things that make them happy.

Red Light: Examining Your Limitations

Life is full of limitations. Time, for example, is always in limited supply, as are natural resources. The second stage of the economic choice model looks at the constraints that force you to choose among your happy options.



For example, oil can be used to manufacture pharmaceuticals that can save many lives. But it can also be used to make petrol, which can be used to drive ambulances, which also save lives. Both pharmaceuticals and petrol are good uses for oil, so society has to come up with some way of deciding how much oil gets to each of these two good uses, knowing all the time that each litre of oil that's put to one use can't be used for the other.

This section outlines the various constraints, as well as the unavoidable cost – the *opportunity cost* – of getting what you want. For more about how markets use supply and demand to allocate resources in the face of constraints, turn to Chapter 8.

Resource constraints

The most obvious constraints on human happiness are the physical limitations of nature. Supplies of oil, water and fish are limited, as are the number of radio frequencies on which to send signals and the hours of sunshine to drive solar-powered cars. There's simply not enough of most natural resources for everyone to have as much as they want.

The limited supply of natural resources is allocated in many different ways. In certain cases – for example, some endangered species – laws guarantee that nobody can have any of the resource. With the electromagnetic spectrum, national governments apportion the spectrum to broadcasters or mobile phone operators. But, for the most part, private property and prices control the allocation of natural resources.

Under such a system, the use of the resource goes to the highest bidder. Although this system can discriminate against the poor because they don't have much to bid with, it does ensure that the limited supply of the resource at least goes to people who value it highly – in other words, to those who have chosen this resource to maximise their happiness.

Technology constraints

No more oil, sunlight or timber exists today than 1,000 years ago, but you have a much higher standard of living than your ancestors. You have a cushier life because of improvements in the technology of converting raw resources into things we like to use.

In just the last 200 years, people have figured out how to immunise children against deadly diseases, how to use electricity to provide light and mechanical power, how to build a rocket capable of putting people on the moon and how to increase farm yields dramatically so that we can feed many more people. In just the last 20 years, the Internet and cheap mobile phones have revolutionised everything from how stock markets price and trade shares to how people receive their electronic entertainment.



As technology improves over time, people are able to produce more from the limited supply of resources on our planet. Or, put slightly differently, as technology improves, we have more and better choices from which to choose.

Yet, because technology improves slowly, how advanced the technology is at any given moment limits our choices. So, thinking of technology as being a constraint that limits choices is quite natural. Fortunately, though, technology does improve over time – meaning that if we just wait, more and better choices become available.

Time constraints

Time is a precious resource. Worse yet, time is a resource in fixed supply. So, the best that technology can do is to allow people to produce more in

the limited amount of time that they have, or to grant a few more years of life through better medical technology.

But even with a longer life span, you can't be in two places at the same time. Otherwise, time wouldn't be a limit and you'd do double the work in the same amount of time. But because you can only be in one place at one time, you're constantly forced to choose, at each and every moment, to do the thing that makes the best possible use of that instant in time.

Opportunity cost: The unavoidable constraint

The economic idea of *opportunity cost* is closely related to the idea of time constraints. You can do only one thing at a time, which means that, inevitably, you're always giving up a bunch of other things.



The opportunity cost of any activity is the value of the next-best alternative thing you may have done instead. For example, imagine you have three options: chatting on the phone with a friend, watching TV or concentrating on reading this chapter. If you choose to chat with your friend because that makes you happiest, we're not going to hold that against you. We just assume that of the two things that you don't choose, you consider reading this chapter to be better than watching TV. So the opportunity cost of chatting on the phone was sacrificing the chance to read this chapter.

Opportunity cost depends only on the value of the next-best alternative. It doesn't matter whether you have 3 alternatives or 3,000. The opportunity cost is simply the value of the next-best alternative because you can always reduce a complicated choice with many options down to a simple choice between two things: option X versus the best alternative out of all the other alternatives.



Opportunity costs can tell you when *not* to do something as well as when to do something. For example, you may love ice cream, but you love chocolate cake even more. If someone offers you only ice cream, you're going to take it. But if you're offered ice cream or chocolate cake, you're going to take the cake. The opportunity cost of eating ice cream is sacrificing the chance to eat chocolate cake. Because the cost of not eating the cake is higher than the benefits of eating the ice cream, it makes no sense to choose ice cream.

Of course, if you choose chocolate cake, you're still faced with the opportunity cost of giving up having ice cream. But you're willing to do that because the ice cream's opportunity cost is lower than the benefits of the chocolate cake. Opportunity costs are unavoidable constraints on behaviour because you always have to decide what's best and give up the next-best alternative.

Making Your Final Choice

At its most basic, the third stage of the economic choice model is nothing more than cost-benefit analysis. In the first stage, you evaluate how happy each of your options is going to make you by measuring how much utility each brings. In the second stage, you determine the constraints and opportunity costs of each option. In the third stage, you simply choose the option for which the benefits outweigh the costs by the largest margin.

The cost-benefit model of how people make decisions is very powerful in that it seems to describe correctly how most decisions are made. But this version of cost-benefit analysis can tell you only whether people choose a given option. In other words, this model is only good at describing all-or-nothing decisions like whether or not to eat ice cream.

A much more powerful version of cost-benefit analysis uses a concept called *marginal utility* to tell you not just whether you're going to eat ice cream, but also *how much* ice cream you're going to decide to eat.

To see how marginal utility works, you need to recognise that the amount of utility that a given thing brings usually depends on how much of that given thing a person has already had. For example, if you've been really hungry, the first slice of pizza that you eat brings you a lot of utility. The second slice is also pleasant, but not quite as good as the first because you're no longer starving. The third, in turn, brings less utility than the second. And if you keep forcing yourself to eat, you may find that the 12th or 13th slice of pizza actually makes you sick and brings you negative utility.

Economists refer to this phenomenon as *diminishing marginal utility*. Each additional, or *marginal*, piece of pizza brings less utility than the previous piece so that the extra utility, or *marginal utility*, brought by each successive slice diminishes as you eat more and more slices.



To see how diminishing marginal utility predicts how people make decisions about how much of something to consume, consider having £10 to spend on slices of pizza or bags of chips. Suppose that slices of pizza cost £2 each, and chips also cost £2 a bag.

Marginal utility is for the birds!

Economists are very confident that cost-benefit analysis and diminishing marginal utility are good descriptions of decision-making because plenty of evidence exists that other species also behave in ways consistent with these concepts.

For example, scientists can train birds to peck at one button in order to earn food and another button to earn time on a treadmill. If scientists increase the cost of one of the options by increasing the number of clicks required to get it, the birds respond rationally by not clicking so much on the button for that option. But even more interesting is that they also switch to clicking more on the button for the other option.

The birds seem to understand that they can make only a limited number of clicks before they get exhausted, and they allocate these clicks between the two options so as to maximise their total utility. Consequently, when the relative costs and benefits of the options change, they change their behaviour quite rationally in response.

Most species also seem to be affected by diminishing marginal utility and become indifferent to the marginal units of something when they've recently enjoyed a large quantity of it. Even bacteria seem to display this behaviour. So although economists' models of human behaviour may seem to ignore some relevant factors, the models do take into account some very fundamental and universal behaviours.

Economists presume that the goal of people faced with a limited budget is to adjust the quantities of each possible thing they can consume to maximise their *total utility*. In this example, because you know that the marginal utility of pizza diminishes quickly with each additional slice, you don't spend all £10 on pizza, because the fifth slice of pizza just isn't going to bring you very much marginal utility. You're better off allocating some of your spending to the chips.

If you buy only four slices of pizza, you free up £2 to spend on a bag of chips. And because the bag of chips is your first bag, eating it probably brings you lots of marginal utility. Indeed, if the marginal utility gained from that first bag of chips exceeds the marginal utility lost by giving up that fifth slice of pizza, you're certain to make the switch. You keep adjusting the quantities of each food until you find the combination that maximises how much total utility you can purchase using your £10.

Because different people have different preferences, the quantities of each good that maximises each person's total utility are usually different. Someone who detests chips spends all his £10 on pizza. A person who can't stand pizza spends all her money on chips. And for people who choose to have some of each, the optimal quantities of each depend on their individual feelings about the two goods and how fast their marginal utilities decrease. Check out

Chapter 9 for more detail on diminishing marginal utility and how it causes demand curves to slope downward.

Allowing for diminishing marginal utility makes this choice behaviour model very powerful: not only does it tell you what people are going to choose, but also how much of each thing they're going to choose. The model's not perfect, however. For example, it assumes that people have a clear sense of the utility of various things, a good idea of how fast marginal utilities diminish and no trouble making comparisons. We discuss these substantial criticisms in the next section.

Exploring Limitations and Violations of the Economist's Choice Model

Economists assume that people are fully informed and totally rational when they make decisions. That's a pretty strong assumption – you'd be justified in asking if it's remotely realistic. The answer to your question is that it varies over the range of decisions we make and the type of behaviour we're looking at. For example, being able to understand risk and probability and take them into account in making our decisions isn't a typical human ability. Other cases might not be remotely representative of a given individual, but after you average them out across everybody in the world the result stands up more robustly. In still others people make decisions that are influenced by the amount of information they have or the behaviour of other people (for instance deciding in which pub to meet up). The model of human behaviour favoured by economists works well as a starting point, but isn't able to tell you with 100 per cent accuracy that Mrs Miggins of East Cheam will choose cream buns over arctic roll on a Tuesday.

Understanding uninformed decision-making

When economists apply the choice model, they assume a situation in which a person knows all the possible options, how much utility each option is going to bring and the opportunity costs of each option. But how do you evaluate whether sitting on top of Mount Everest for five minutes is better than hang-gliding over the Amazon for ten minutes? If you've never had either experience, you aren't well-informed about the constraints and costs of the choice and probably don't even know the utilities of the two options.

Politicians with novel new programmes often ask people to make similarly uninformed choices. They make their proposals sound as good as possible, but in many cases nobody really knows what they may be getting into.

Things are similarly murky when making choices about random events. People buying lottery tickets in the national lottery have no idea about the eventual possible gain or the eventual likelihood of winning, because both the size of the prize and the likelihood of winning depend on how many tickets may or may not be sold before the draw is made.

Economists account for this reality by assuming that when faced with uninformed decisions, people make their best guesses about not only random outcomes, but also about how much they may like or dislike things with which they have no previous experience. Although this may seem like a fudge, because people in the real world are obviously making decisions in such situations (they do, in fact, buy a whole lot of lottery tickets), the people in those situations must be fudging a bit as well.

Whether people make good choices when they are uninformed is hard to say. Obviously, people prefer to be better informed before choosing. And some people do shy away from less certain options. But, overall, the economist's model of choice behaviour seems quite capable of dealing with situations of incomplete information and uncertainty about random outcomes.

Getting rational about irrationality

Even when people are fully informed about their options, they often make logical errors in evaluating the costs and benefits of each. We go through three of the most common choice errors in the following sections, but as you read them, don't be too alarmed. After these logical errors are explained to them, people typically stop making the errors and start behaving in a manner consistent with rationally weighing marginal benefits against marginal costs.

Sunk costs are sunk!

Suppose that you just spent £15 to get into an all-you-can-eat sushi restaurant. How much should you eat? More specifically, when deciding how much to eat, should you care about how much you paid to get into the restaurant?

To an economist, the answer to the first question is: eat exactly the amount of food that makes you most happy. And the answer to the second question is: how much it costs you to get in doesn't matter because whether you eat one piece of sushi or 80 pieces of sushi, the cost is the same. Put differently, because the cost of getting into the restaurant is now in the past, it should be completely unrelated to your current decision of how much to eat.



Economists refer to costs that have already been incurred and which should therefore not affect your current and future decision-making as *sunk costs*. Rationally speaking, you should consider only the future, potential marginal costs and benefits of your current options.

After all, if you were suddenly offered £1,000 to leave the sushi restaurant and eat next door at a competitor, would you refuse simply because you felt you had to eat a lot at the sushi restaurant in order to get your money's worth out of the £15 you spent? Of course not.

Unfortunately, most people tend to let sunk costs affect their decision-making until an economist points out to them that sunk costs are irrelevant, or, as economists never tire of saying, 'Sunk costs are sunk!' (On the other hand, non-economists quickly tire of hearing this phrase.) Fortunately, we have other ways of saying the same thing: for example, we talk about the Concorde fallacy, in honour of the supersonic jet that never broke even, no matter how many pounds or francs were thrown at the project.

Mistaking a big percentage for a big pound amount



Suppose you decide to save 10 per cent on a TV by making a one-hour round trip to a store in another town to buy the TV for only £90 instead of buying the TV at your local store for £100. Next, ask yourself whether you'd also be willing to drive one hour in order to buy a home theatre system for £1,990 in the next town rather than for £2,000 at your local store. You do the maths, and because you're going to save only 0.5 per cent, you decide to buy the system for £2,000 at the local store.

You may think you're being smart, but you've just behaved in a colossally inconsistent and irrational way. In the first case, you were willing to drive one hour to save £10. In the second, you were not. Costs and benefits are absolute, but people make the mistake of thinking of the costs and benefits of driving to the next town in terms of percentages or proportions. Instead, compare the total costs against the total benefits because the benefit of driving to the next town is the absolute amount in pounds you save, not the proportion you save.

Confusing marginal and average



Suppose that your local government recently built three bridges at a total cost of £30 million: that's an average cost of £10 million per bridge. A local economist does a study and estimates that the total benefit of the three bridges to the local economy adds up to £36 million, or an average of £12 million per bridge.

A politician then starts trying to build a fourth bridge, arguing that because bridges on average cost £10 million but on average bring £12 million in benefits, it would be foolish not to build another bridge. Should you believe him? After all, if each bridge brings society a net gain of £2 million, you'd want to keep building bridges forever.

However, what really matters to this decision are *marginal* costs and *marginal* benefits, not *average* ones (see the section 'Making Your Final Choice' for more on marginal utility). Who cares what costs and benefits all the previous bridges brought with them? You have to compare the costs of that extra, marginal bridge with the benefits of that extra, marginal bridge. If the marginal benefits exceed the marginal costs, you should build the bridge. And if the marginal costs exceed the marginal benefits, you shouldn't.

For example, suppose that an independent watchdog group hires an engineer to estimate the cost of building one more bridge and an economist to estimate the benefits of building one more bridge. The engineer finds that because the first three bridges have already taken the three shortest river crossings, the fourth bridge needs to be much longer. In fact, the extra length raises the building cost to £15 million.

At the same time, the economist does a survey and finds that a fourth bridge isn't really all that necessary. At best, the bridge is going to bring only £8 million per year in benefits. Consequently, this fourth bridge shouldn't be built because its marginal cost of £15 million exceeds its marginal benefit of £8 million. By telling voters only about the *average* costs and benefits of past bridges, the politician supporting the project is grossly misleading them. So watch out whenever somebody tries to sell you a bridge.

Chapter 3

Producing the Right Stuff in the Right Way to Maximise Human Happiness

In This Chapter

- ▶ Determining your production possibilities
 - ▶ Allocating resources in the face of diminishing returns
 - ▶ Choosing outputs that maximise people's happiness
 - ▶ Understanding the role of government and markets in production and distribution
-

An optimist sees half a glass of milk and says the glass is half full. A pessimist sees half a glass of milk and says the glass is half empty. An economist sees half a glass of milk and wonders why the glass is twice as big as it needs to be.

Anonymous

Although human beings face scarcity and can't have everything they want (as we discuss in Chapter 2), they do have a lot of options. Productive technology is now so advanced that people can convert the planet's limited supply of resources into an amazing variety of goods and services, including cars, computers, aeroplanes, cancer treatments, video games and even totally awesome *For Dummies* books like this one.

In fact, thanks to advanced technologies, people (in developed economies, at least) are spoiled for choice. The huge variety of goods and services that can be produced means that people must choose wisely if they want to convert the planet's limited resources into the goods and services that are going to provide the greatest possible happiness when consumed.

This chapter explains how economists analyse the process by which societies choose exactly what to produce in order to maximise human happiness. For every society, the process can be divided into two simple steps:

1. **The society must figure out all the possible combinations of goods and services that it can produce given its limited resources and the currently available technology.**
2. **The society must choose one of these possible output combinations – presumably, the combination that maximises happiness.**

Economists view success in each of the two steps in terms of two particular types of efficiency:

- ✓ *Productive efficiency* means producing any given good or service using the fewest possible resources.
- ✓ *Allocative efficiency* means producing the kinds of goods and services that are going to make people most happy, and producing them in the correct amounts.

This chapter shows you how a society achieves both productive and allocative efficiency – that is, how a society determines what's possible to produce, as well as what's best to produce. Here, we give you the basics on diminishing returns, Production Possibilities Frontier graphs and the interplay between markets and governments.

Reaching the Limit: Determining What's Possible to Produce

In determining what's possible to produce in an economy, economists list two major factors that affect both the maximum amounts and the types of output to be produced:

- ✓ Limited resources
- ✓ Diminishing returns



The first factor is obvious: if resources were unlimited, goods and services would be as well. The second factor, despite affecting nearly every production process known, isn't understood by most people. Basically, *diminishing returns* means that the more you make of something, the less return you get on each successive unit. Eventually, the costs exceed the benefits, which limits how much of it you want to produce, even if the product is your favourite thing. Your resources should be devoted to producing units of other things for

which the benefits still outweigh the costs. Diminishing returns implies that, in general, we're better off not putting all our eggs in one basket.

This section gives you the low-down on how limited resources and diminishing returns determine production possibilities. We also show you how to represent these possibilities graphically.

Classifying resources used in production

You can't get output without inputs of resources. Economists traditionally divide inputs, or *factors of production*, into three classes:

- ✔ **Land:** To economists, land means a little more than just real estate or property. Land also refers to all naturally occurring resources that can be used to produce things people want to consume. Land includes the weather, plant and animal life, geothermal energy and the electromagnetic spectrum.
- ✔ **Labour:** The work that people must do in order to produce things. A tree doesn't become a house without human intervention.
- ✔ **Capital:** Man-made machines, tools and structures that aren't directly consumed but are used to produce other things that people do directly consume. For example, a car that you drive for pleasure is a consumption good, whereas an identical car that you use to haul around bricks for your construction business is capital. Capital includes factories, roads, sewers, electrical grids, the Internet and so on.

In addition to these three traditional inputs, economists now often speak of *human capital*, which is the knowledge and skills that people use to help them produce output. For example, we each have a lot of human capital with regard to teaching economics, but we have extremely low human capital with regard to painting and singing. (Be very happy that you haven't heard us sing!)

If you put a person to work at a job for which he or she has high human capital, he or she produces much better or much more output than a person with low human capital, even though they both supply the same amount of labour in terms of hours worked. An important consequence is that skilled workers (high human capital) get paid more than unskilled workers (low human capital). Therefore, a good way for societies to become richer is to improve the skills of their workers through education and training. If societies can raise workers' human capital levels, not only can they produce more with the same inputs of limited land, labour and capital, but also their workers are going to be paid more and enjoy higher standards of living.

An easy way to think about the idea of human capital is to compare two builders who have equal access to tools (capital) and work the same amount of time (labour), but have different levels of human capital, perhaps due to one having better training than the other. The one with the higher level of human capital can be more productive (able to do more jobs in the same amount of time, or even having fewer bungles and bodes to fix) than the one with the lower level of human capital.

But building up human capital is costly, and at any given instant, you need to think of the level of human capital in a society as being fixed. Combined with limitations on the amount of land, labour and capital, the limitation on human capital means that the society can only produce a limited amount of output. And along these same lines, the decisions about where to best allocate these limited resources become crucial because the resources must be used for production of the goods and services that are going to bring with them the greatest amount of happiness. (For more about limited resources and production possibilities, see the upcoming section ‘Allocating resources: A little here, a little there’.)

Getting less of a good thing: Diminishing returns



Diminishing returns is probably the most important economic factor in determining exactly what to produce out of all the things that can possibly be produced given the limited supply of resources. Diminishing returns refers to the fact that for virtually everything people make, the amount of additional output you get from each additional unit of input decreases as you use more and more of the input.

In our discussion of diminishing returns, do bear in mind that we are discussing returns to one changing factor whilst holding others constant. This is key. It makes sense to talk about diminishing returns to labour holding capital fixed. If capital can change too, we’d have to adapt the model by making a new calculation for returns to labour every time we changed the amount of capital used. We get around that by holding capital constant, that is, working out return to labour *for a given amount of capital used*.



Diminishing returns is sometimes referred to as the *low-hanging fruit principle*. Imagine being sent into an apple orchard at harvest time to pick apples. During the first hour, you pick a lot of apples because you go for the low-hanging ones that are the easiest to reach. In the second hour, however, you can’t pick as many because you have to start reaching awkwardly for fruit that is higher up. During the third hour you pick even fewer apples; you now have to jump off the ground every time you try to pick an apple because the only ones left are even farther away. Table 3-1 demonstrates how your productivity – your output for a given amount of input – diminishes with each additional hour you work.

<i>Hour Worked</i>	<i>Apples Picked</i>	<i>Labour Cost per Apple</i>
1	300	2p
2	200	3p
3	120	5p

Another way to see the effect of diminishing returns is to note the increasing costs for producing output. If you pay workers £6 per hour to pick apples, your cost to have 300 apples picked in the first hour is 2 pence per apple, as shown in Table 3-1. The second hour yields only 200 apples, costing you 3 pence per apple (because you still have to pay the worker £6 for that hour's work). Only 120 apples get picked in the third hour, so the labour cost per apple rises to 5 pence.

Eventually, the effects of diminishing returns drive prices so high that you stop devoting further labour resources to picking additional apples.

Virtually all production processes show diminishing returns, and not just for labour. Additional amounts of any particular input usually result in smaller and smaller increments of output, holding all other inputs constant. (In some special cases this doesn't hold or it can look for a while as if it doesn't, but this is pretty much the case for almost all production.)

Allocating resources: A little here, a little there

Because the diminishing returns factor assures that a production process eventually becomes too costly, a society normally allocates its limited resources widely, to many different production processes.



To understand why this happens, imagine that you can allocate workers to picking apples or picking oranges and crucially you have a fixed number of ladders. You can sell both apples and oranges for £1 each, but the production of both fruits involves diminishing returns so that additional workers acting as fruit pickers yield successively smaller increases in output no matter which fruit they're picking.

Allocating all your workers to picking oranges, for example, is unproductive because the output you get from the last worker picking oranges is much less than the output you get from the first worker picking oranges.

The smart thing to do is to take a worker away from picking oranges and reassign him to picking apples. As the last worker picking oranges, he didn't produce much. But as the first worker picking apples, he's going to pick a lot of them. Because you pay him the same wage regardless of which fruit he picks, you use your labour more intelligently by having him pick apples, because one apple sells for as much money as one orange.

You may also want to reassign a second worker, and perhaps a third or a fourth. But because diminishing returns applies just as much to picking apples as it does to picking oranges, you don't want to reassign all the workers. Each additional worker assigned to picking apples produces less than the previous worker picking apples. At some point, moving additional workers from picking oranges to picking apples no longer benefits you, and you've reached what economists refer to as an *optimal allocation* of your labour resource. As soon as you've found this sweet spot, you have no further incentive to move workers from picking one fruit to picking the other because no additional moving of workers increases total fruit picking. At this point, you've maximised your fruit-picking potential.

Graphing your production possibilities

Economists have a handy graph called the *Production Possibilities Frontier* (PPF) that lets you visualise the effect of diminishing returns and view the trade-offs you make when you reallocate inputs from producing one thing to producing another. The Production Possibilities Frontier, which is sometimes referred to as the *Production Possibilities Curve*, also shows how limited resources limit your ability to produce output. Figure 3-1 shows a PPF graph that corresponds to the data in Table 3-2.

Table 3-2 shows how the total output of apples and oranges changes as you make different allocations of five available workers to picking apples or oranges. For example, if you put all five people to work picking only apples for one whole day, you get 700 apples picked and zero oranges picked. If you move one worker to oranges (so four workers are picking apples and one worker is picking oranges), you get 680 apples picked and 300 oranges picked. Because of diminishing returns, taking one worker away from apples reduces apple output by only 20. But moving that worker to oranges increases orange production by 300 because that worker is the first one picking oranges and can get the low-hanging fruit.

	<i>Combo 1</i>	<i>Combo 2</i>	<i>Combo 3</i>	<i>Combo 4</i>	<i>Combo 5</i>	<i>Combo 6</i>
Workers picking oranges	0	1	2	3	4	5
Workers picking apples	5	4	3	2	1	0
Output of oranges	0	300	500	620	680	700
Output of apples	700	680	620	500	300	0

Figure 3-1 plots out the six output combinations that result from varying the allocation of workers in Table 3-2, thereby graphing all your production possibilities. Point A corresponds to putting all your workers to work picking apples. Point B corresponds to the output you get from four workers picking apples and one worker picking oranges.

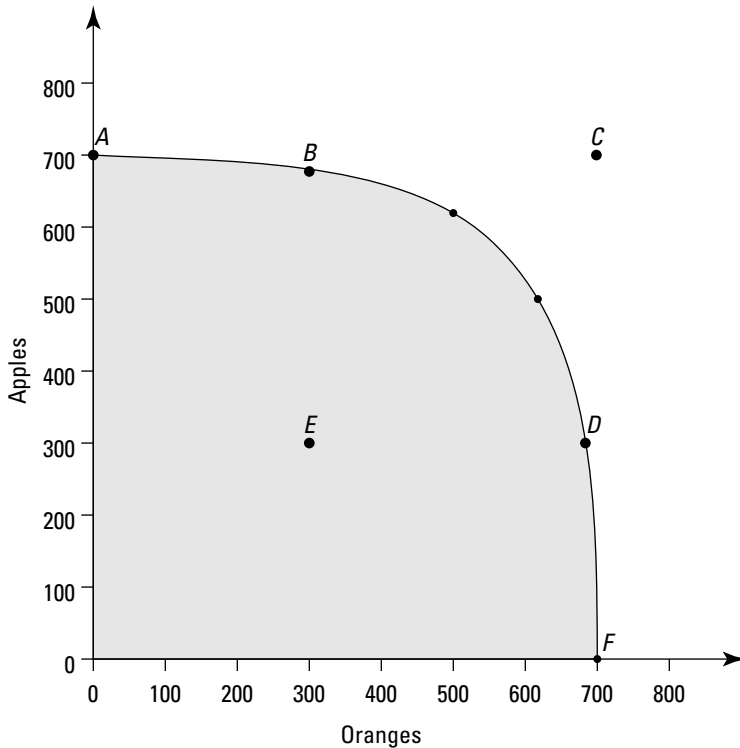


Figure 3-1:
The
Production
Possibilities
Frontier
(PPF) for
the data in
Table 3-2.

Note that each of the six points is *attainable* in the sense that you can actually produce the corresponding quantities of each fruit through some allocation of the five workers' labour. On the other hand, a point like *C* is not attainable. You can't allocate your five workers in any way to produce that many apples and oranges. Perhaps if you had more workers you'd be able to produce such an output combination, but you're limited with only five workers.

Imagine that instead of allocating labour by worker, you allocate it by time. The five workers each work for one day, so you have five *worker-days* of labour to allocate. You can now allocate, for example, 3.2 worker-days to apple picking and 1.8 worker-days to orange picking. This arrangement allows you to fill in the graph and draw a line connecting the six points that correspond to the output combinations that you get when allocating labour by worker.

This line is called the *Production Possibilities Frontier*, or PPF, because it divides the area of the graph into two parts: the combinations of output that are possible to produce given your limited supply of labour are under the line, and those that are not possible to produce are above it. In this way, the PPF graph captures the effect of scarce resources on production. Some output combinations are just not producible given the limited supply of labour.

The changing slope as you move along the frontier shows that the trade-off between apple production and orange production depends on where you start. If you're at point *A*, where you're allocating all your resources to the production of apples, you can, by reallocating resources, produce a lot more oranges at the cost of giving up only a few apples. But if you start at point *D*, where you're already producing a lot of oranges, you have to give up a lot of apples to get just a few more oranges.

In economic jargon, the changing slope of the PPF in the face of diminishing returns is due to the fact that the *opportunity costs* of production vary depending on your current allocation of resources. (Check out Chapter 2 for more on opportunity costs.) If you're already producing a lot of apples, the opportunity costs of devoting even more labour to more apple production are very high because you're giving up a lot of potential orange production. On the other hand, the opportunity costs of devoting that labour to orange production are very low because you have to give up producing only a few apples. Clearly, you should devote the labour to picking the fruit that has the lower opportunity costs because, in this example, both fruits bring the same benefit: £1 per fruit sold.



The PPF is also very handy because any points that lie on the PPF itself (on the frontier) clearly show the output combinations you get when you're *productively efficient*, or wasting none of your resources. You can't increase the production of apples without reducing the production of oranges, and vice versa. For example, if you start at point *B*, the only way to increase apple production is to slide up along the frontier, which implies reducing orange production. You have to make this trade-off because you don't have any wasted labour lying around with which you can get more apples without reducing the amount of labour already devoted to orange picking.

All the points below the line are productively inefficient. Consider point *E* in Figure 3-1, which corresponds to producing 300 apples and 300 oranges. You produce at a point like *E* only if you're being productively inefficient. In fact, you can see from Table 3-2 that you can produce these numbers by sending only one worker to pick apples and another worker to pick oranges. You're using just two of your five workers; the labour of the other three workers is being wasted or not used at all. (This illustration also illuminates the punch line of our joke at the beginning of this chapter – an economist may think the glass is twice as big as it needs to be, in other words, that the glass exhibits productive inefficiency!)

In the real world, you end up at points like *E* because of inefficient production technology or poor management. For one reason or another, the resources that are available aren't being used to produce as much output as possible.

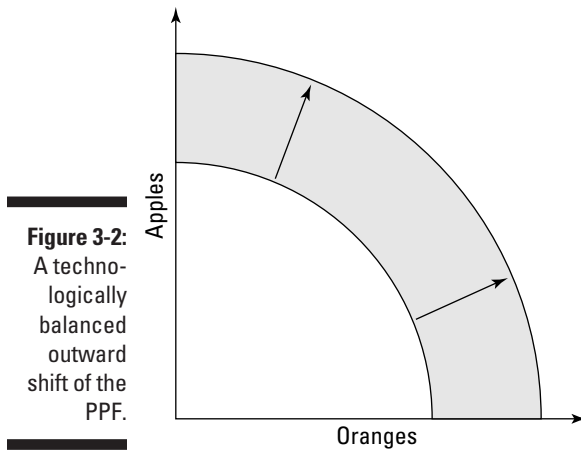
Any manager who has five workers to allocate but produces only output combination *E* would be fired! If we were to aggregate across all the firms in an economy, we'd say that efficient economies should always be producing at some point on their frontiers because if they are inside that frontier line, they are wasting their limited resources and not maximising the happiness that can be had from them.

Pushing the line with better technology

This PPF is a simplification of the real world, derived by allocating one input between just two outputs. The real world is, of course, more complicated, with many different resources allocated among many different outputs. But the principles of limited resources and diminishing returns that show up so clearly on the PPF graph also apply to the much greater variety of both inputs and outputs in the real world.

Another simplification of the diminishing returns model is that, other than the particular input you are allocating, you are implicitly holding constant all other productive inputs, including technology. But humanity's level of technological sophistication is constantly increasing, allowing people to produce much more from a given set of resources than before. So whilst this is a neat model for getting a handle on the problem, you have to adapt it to use in the real world.

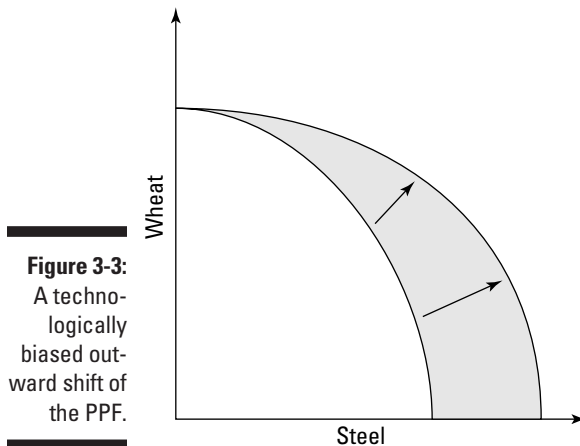
Economists represent this increase in productivity by shifting the PPF outward. In Figure 3-2, the shaded area represents new combinations of output that, thanks to better technology, can now be produced using the same amount of resources as before. The PPF is still curved because better technologies don't get rid of diminishing returns. Even with a better technology, if you start increasing the amount of a particular input, you get successively smaller additional increases in output.



In Figure 3-2, the new technology shift is *balanced* in the sense that it increases your ability to produce more of both goods. An example of a balanced technological change is improvements in fertilisers or pesticides that increase crop yields of both apples and oranges.

But most technological innovations are *biased*. For example, suppose that you're considering a PPF where the two output goods are wheat and steel. An improvement in steel-making technology obviously allows you to make more steel from your limited resources but has no effect at all on your ability

to make wheat. Consequently, as Figure 3-3 shows, the PPF doesn't shift out evenly. Rather, it shifts out at the end where all your particular input (say, labour) is devoted to steel, but remains fixed at the end where all your particular input is devoted to wheat production.



Determining What Should Be Produced

After a society locates the frontier of efficient output combinations, the next step is choosing the point along the frontier that produces the combination of goods and services that makes people most happy. Choosing only from among frontier combinations guarantees *productive efficiency*. Choosing the single frontier combination that maximises happiness assures *allocative efficiency*.

Because determining where the frontier lies is mostly a matter of engineering and applying current technology to available resources, it engenders little controversy. But deciding which particular combination of outputs a society as a whole should choose is much more complicated. People have preferences both as individuals and as groups about what products make them happiest. An individual choosing a point along his own personal PPF encounters no conflict. He just determines what combination of output makes him happiest and then he produces and consumes it.



The decision-making process becomes vastly more complicated when you consider an entire society's PPF, in which case you're sure to have vigorous disagreement about what combination of output to produce with the society's limited resources. For example, your neighbour may not mind all the pollution produced by the fact that he likes driving his 4x4 vehicle day and night. If he were living in his own world, the pollution wouldn't matter, but because you live near him, you're affected by the pollution and so you object. Perhaps you seek government intervention to limit what your neighbour is doing. Similarly, the government argues over what it should produce with its limited resources: some people favour farm subsidies, whereas others favour defence spending or programmes to aid the poor.

These competing priorities mean that some sort of decision-making process must be established to determine what actually gets produced and to (try to) make sure that it pleases most of the people most of the time.

In most modern economies, this process is the result of both private and public decisions acting through a combination of free markets and government action. The process is not always smooth – and budget negotiations have driven plenty of politicians to drink or despair – but it has delivered the highest living standards in world history.

Weighing pros and cons of markets and government interventions

When analysing the ways in which modern economies and societies select a combination of goods and services to produce, you have to realise that current economic laws and institutions are the result of conflicting pressures about whether to leave markets to their own devices when turning resources into output or use the power of government to intervene in markets in order to secure a different set of outcomes.

Keep the following three factors in mind when considering the fight between leaving the markets alone and intervening:

- ✓ Modern economies are hugely complicated, with literally millions of goods and services produced using limited supplies of land, labour and capital. Markets handle this complexity easily, but government interventions usually don't – meaning that they often risk substantial reductions in productive and allocative efficiency.

- ✔ Some goods and services, such as coal-burning power plants and cocaine, have negative consequences. These negative consequences bring forth substantial pressure for government intervention in the economy because these markets, if left alone, produce a lot of these goods and services.
- ✔ Some people end up consuming a very large proportion of the goods and services produced, whereas others end up with very little. Such unequal distribution also brings forth a great deal of pressure for government intervention in the economy in order to equalise living standards.

These factors are both a consequence and a cause of the fact that our modern economies are largely a mix of market production and government intervention. For the most part, what to produce, how much of it to produce and who gets it is decided by voluntary transactions made by individuals and businesses. But sometimes, the government uses its coercive powers to achieve outcomes that wouldn't happen if individuals and businesses were left to their own devices.

In both cases, a huge apparatus of law and tradition governing economic transactions helps society produce a combination of output that is, hopefully, both productively efficient (so resources aren't wasted) and allocatively efficient (so the economy is producing the things that people want most). Next, we outline the benefits and the drawbacks that both markets and governments bring to the economic table.

The magic of markets: Going where no one person can ever go

Market production is the term that economists use to capture what happens when one individual offers to make or sell something to another individual at a price agreeable to both. Markets are very good at producing things for which people are willing to pay. In addition, markets tend to be very efficient if many providers of a good or service exist.

A *competitive market* is one in which many sellers compete against each other to attract customers. In such a situation, each seller has an incentive to sell at the lowest price possible subject to constraints imposed by their costs in order to undercut competitors and steal their customers. Because every firm has this incentive, prices tend to be driven so low that the businesses can just barely make a profit.



A competitive market also tends to guarantee productive efficiency because the best way for sellers to keep prices low is to make sure that they're using all their resources efficiently and that nothing is going to waste. Because competition is ongoing, the pressure to be efficient is constant. Sellers also have a big incentive to improve efficiency in order to undersell their rivals and steal their customers.

In terms of the PPF (which we discuss in the earlier section 'Graphing your production possibilities'), market production with a lot of competition tends to ensure not only that economies produce along the frontier, but also that they have frontiers that are constantly being pushed outward as firms improve efficiency.

Markets also have the benefit of working out, automatically, the things that people want. To grasp why this is so amazing, consider that we live in a world of nearly seven billion people. It would be very hard for any one person to gather enough information to figure out what each of those nearly seven billion people most wants to buy. Several lifetimes would be needed to speak with each of them, even just to find out what they want for dinner, let alone all the other things they may most like to purchase on a typical day.

But because production and distribution in modern economies aren't centralised, you don't need to know the big picture. In fact, the real magic of market economies is that they are just a collection of millions and billions of small face-to-face transactions between buyers and sellers.



For example, the person who sells you a TV at the local store has no idea about the total demand for TVs in the world, how many tons of steel or plastic are needed to produce them, or how many other things *weren't* produced because the steel and plastic needed to make the TVs were used for TVs rather than other things. All they know is that you're willing to pay them for a TV. And if they're making a profit selling TVs, they will take account of the greater potential sales by ordering more TVs from the factory. The factory, in turn, increases production, taking resources away from the production of other things. Reallocation of resources also occurs in markets because each resource has a price, and whoever is willing to pay the price gets the resource.



In fact, market economies are often called *price systems* because prices serve as the signals that allocate resources. Things in high demand have high prices, and things in low demand have low prices. Because businesses like to make money, they follow the price signals and produce more of what has a high price and less of what has a low price. In this way, markets tend to take our limited resources and use them to produce what people most want – or, at least, what people are most willing to pay for. And they do it all in a completely decentralised manner.



Communism, long lines and toilet paper

In a *command economy* all economic activity is done on the orders of the government. Until the fall of the Berlin Wall and the subsequent collapse of communism in the late 1980s and early 1990s, a large part of the world's population lived in countries that had command economies. Sadly, they didn't live very well.

Shortages of everything from sugar to clothing to toilet paper were constant. More seriously, doctors often lacked hypodermic needles and medicines for their patients, and food was often in short supply.

Goods and services weren't allocated using a price system in which output went to those willing and able to pay for it. Instead, because everyone in a communist country is ideologically equal, the government attempted to give everyone an equal share of the goods and services made. The result, though, wasn't an equal division; instead, there were long lines, with those able to stand in line the longest getting more than their fair share. The lines were so long that people often stood in line for an entire day just to get one roll of toilet paper. If you saw a line forming, you got in it as fast as possible, even if you didn't know what people were standing in line for. Because everything was in short supply, the product was almost certainly something you'd want.

What caused this mess? Centralisation. In Moscow, government officials called *central planners* attempted to determine the correct amounts to produce for 24 million different items! It was an impossible task. Take, for example, toilet paper. First, you estimate how many millions of rolls of toilet paper are needed. Then you have to figure out how many trees to cut down to make that much paper and how many railcars you need to carry those trees to paper mills and how many workers it takes to run those mills. At the same time, you have to try to balance production of toilet paper against the other zillion things that also require trees, railcars and workers.

The entire problem is far too complex and requires far too much information to be solved. The result was that resources were constantly being misdirected and wasted. For example, food often rotted at farms because no railcars had been scheduled to take it to cities; the officials hadn't accounted for an early harvest, and the railcars were busy elsewhere. In a price system, the farmers would have simply paid to bid the railcars away from other uses. This solution wasn't possible in a centralised economy in which prices weren't used to allocate resources.

The misdeeds of markets

Markets aren't perfect. In particular, they suffer from two major problems:

- ✔ Markets produce whatever people are willing to pay for, even if these things aren't necessarily good for the people or the environment.
- ✔ Markets are amoral: they don't in any way guarantee fairness or equity.



The fact that illegal drugs are widely and cheaply available despite vigorous government programmes to stop their production and distribution is probably the best example of the robustness of markets. As long as there are profits to be made, you can be pretty certain that supply is going to rise to satisfy any demand. Having markets that are hell-bent on giving people what they're willing to pay for is all very well, but illegal drugs are an excellent example of the fact that markets can deliver things without caring about their social value or their negative consequences.

Along the same lines, producers often do things we don't like while giving us what we want. Child labour and sweatshop labour are prime examples. Often the government is called on to intervene when the price system doesn't provide enough incentive for producers to change such objectionable practices.

The other big problem with markets is that they cater to those who have money to spend. The price system gives an incentive to produce only the things that people are willing and able to pay for. If someone is very poor, he can't give producers an incentive to provide him with even basic necessities like medicine and food. Under a pure price system, resources are instead directed toward producing things for those who have money to spend.

A related problem with markets is income and wealth inequality. Because market systems reward those who are best able to provide goods and services that people want to buy, some sellers end up becoming very rich because they're better at providing what people want. This invariably leads to large inequalities in wealth that many people find offensive, even when the money is honestly earned and even though highly productive people make such large contributions toward increasing output and maximising happiness.

The case for government intervention

Many societies use their governments to intervene and address the problems that markets create or can't fix. Government interventions in the economy usually take one of three forms:

- ✓ **Penalties or bans on producing or consuming goods or services that are considered dangerous or immoral:** For example, governments may ban drugs or impose 'sin taxes' on things like alcohol and tobacco, which, though legal, are thought to be products whose use should be discouraged. However, prohibitions seldom work because the market, which need not be a *legal* market, still has large incentives to provide such goods and services. Taxes, while smarter, work up until the tax on the product is high enough that an illegal market can undercut the taxed market.

- ✔ **Subsidies to encourage the production of goods and services that are considered desirable:** For example, most governments heavily subsidise the education of children and the provision of medical care. They do so because of the fear that insufficient education and inadequate medical care are likely to be provided without these subsidies.
- ✔ **Taxes on the well-off to provide goods and services to the less fortunate and to reduce inequalities in income and wealth:** These taxes are put toward things like good parks, clean air and art, as well as goods and services for the poor. Governments tax individuals and businesses in order to raise the money to provide such things.

In terms of the PPF graph, each of these government interventions causes the economy to produce and allocate an output combination different from the one that society would end up with if the markets made all the production and allocation decisions.

Depending on the situation, the output combination produced by a government intervention may be better or worse than the market combination in terms of productive efficiency, allocative efficiency or both. The best combination depends on the specifics of each case.

The case against government intervention

Government intervention is a powerful force for redirecting economic activity, but it doesn't necessarily make the economy better. In fact, we can think of at least three good reasons why government interventions in the economy make things worse:

- ✔ Government programmes are often the result of special-interest lobbying that seeks to help some small group rather than to maximise the happiness of the general population. Special-interest lobbying takes resources away from other uses that often benefit numerous people in order to provide benefits to only a few.
- ✔ Government programmes often deliver poor service even when pursuing the common good, because they have no competition to create incentives to produce government goods and services efficiently.
- ✔ Government interventions usually lack the flexibility of the price system, which is able to constantly redirect resources to accommodate people's changing willingness to pay for one good rather than another. Government policies take years to pass, and laws are usually written in a very precise manner that doesn't allow for changing circumstances and rapid innovation – things that the price system handles with ease.



The wine lakes

In the years immediately following the Second World War, European policymakers were keen to learn lessons from the past. One of those lessons was that falling farm prices were a source of impoverishment for many of the small-landholding farmers that made up the bulk of European agriculture. For many reasons, policymakers were keen never to replicate the economic conditions of the interwar years, so they thought of a solution to the problem.

The solution they came up with was simple in theory but had some practical consequences that were probably unintended. They developed what is known as the *Common Agricultural Policy (CAP)*. Put simply, this policy would guarantee a minimum price for farmers and small-landholders, with the revenue made up from general taxation (via value added tax, or VAT).

The problem was that technical progress and greater efficiency in agriculture combined with the guaranteed price to give an incentive to overproduction. Because farmers could claim subsidies simply for producing goods that nobody actually needed to buy, they always had an incentive to produce more than the market needed. As a result of the overproduction, European consumers paid higher prices (through guaranteed pricing), higher taxes (for the subsidy to overproduce) and got lakes of wine that no one would ever drink and mountains of butter that no one would ever eat. Meanwhile, reforming the CAP is now regarded as one of the most intractable problems in European politics as well as one of the most pressing.

Although markets sometimes fail to deliver everything that society wants, government intervention isn't a panacea. Markets are very good at delivering the vast majority of things that people want, and they can usually do so at the lowest possible cost. Consequently, government intervention needs to be well thought out in case it makes things worse rather than better.

Opting for a mixed economy

In the real world, few societies opt for an extreme type of economy, such as one that is totally market-based or one that features constant and pervasive government intervention. Instead, most societies opt for some mixture of markets, government intervention and what economists refer to as *traditional production*. In their purest forms, these three types of economy can be defined as follows:

- ✔ **A market economy** is one in which almost all economic activity happens in markets with little or no interference by the government. Because of the lack of government intervention, this system is also often referred to as *laissez-faire*, which is French for 'let well alone'.
- ✔ **A command economy** is one in which all economic activity is directed by the government.
- ✔ **A traditional economy** is one in which production and distribution are handled along the lines of long-standing cultural traditions. For example, until the caste system was abolished in India during the last century, the production of nearly every good and service was permitted only by someone born into the appropriate caste. Similarly, in medieval Europe, people were usually unable to be part of the government or attain high military rank unless they were born into the nobility.

Because nearly every modern economy is a mixture of these three pure forms, most modern economies fall into the very inclusive category called *mixed economies*. With the exception of a few isolated traditional societies, however, the traditional economy part of the mixture has tended to decline in significance because most production has shifted to markets and because traditional economic restrictions on things like age and gender have become less important (and more illegal).



The result is that most mixed economies today are a mixture of the other two pure types: the command economy and the market economy. The mixtures that you find in most countries typically feature governments that mostly allow markets to determine what's produced, but that also mix in limited interventions in an attempt to make improvements over what the market would do if left to its own devices.

The precise nature of the mixture depends on the country, with the United Kingdom and the United States featuring more emphasis on markets whereas France and Germany, for example, feature more emphasis on government intervention. On the other hand, a few totalitarian states like North Korea still persist in running pure command economies as part of their all-encompassing authoritarian regimes. However, even among totalitarian states, pure command systems are rare. For example, even in Cuba a command economy exists alongside a thriving small business economy, and other one-party states such as China or Vietnam have made great strides in freeing the business sector of their economies.

As we discuss in the sidebar ‘Communism, long lines and toilet paper’ in this chapter, command economies have all been dismal failures when we evaluate them in terms of productive and allocative efficiency, although plenty of people still feel nostalgia for them. Even well-intentioned governments can’t gather enough information about production and distribution to do a good job of allocating resources. In fact, they do a much worse job than price systems.

Consequently, the opposite extreme, absolutely no government intervention, is an attractive option. Such laissez-faire systems were first suggested by French economists a couple of centuries ago in response to the habit of governments of that era to intervene very heavily in economic activity.

However, no pure laissez-faire economy has ever existed or is probably ever going to exist. The simple fact is that properly functioning market economies that use price mechanisms to allocate resources require a huge amount of government support. Among other things, market economies need governments to do the following:

- ✔ Enforce property rights so that people don’t steal
- ✔ Provide legal systems to write and enforce contracts so that people can make purchases and sales of goods and services
- ✔ Enforce standardised systems of weights and measures so that people know they aren’t being cheated
- ✔ Provide a stable money supply that’s safe from counterfeiters
- ✔ Enforce patents and copyrights to encourage innovation and creativity

Notice that all these things must be in place in order for markets to function. Consequently, a more moderate, more modern version of laissez-faire says that government should provide the institutional framework necessary for market economies to function, and then get out of the way and let people make and sell whatever is demanded.

However, the vast majority of people want governments to do more than just set up the institutions necessary for markets to function. They want governments to stop the production and sale of things like drugs or subsidise the production of things that the market economy may not provide a lot of, such as housing for the poor. They often also want to tax well-off citizens to pay for government programmes for the poor.



Many government programmes are so commonplace that you don't even think of them as being government interventions. For example, free public schools, safety features on cars, warning labels on medicine bottles, sin taxes on alcohol and tobacco and mandatory contributions to retirement systems are all government interventions in the economy.

The government interventions needed to implement such programmes are, in some cases, not efficient. But many people argue that life consists of, and requires, more than efficiency, and therefore the inefficiencies caused by many government interventions are well worth the benefits that they produce. For such people, the government interventions in question increase overall happiness despite the fact that they are, strictly speaking, inefficient.

Because pure market economies don't deliver everything that many people want, most societies have opted for at least some – and in some cases, quite a lot of – government intervention in the economy. The result is that most economies today are mixed economies, with some aspects of direct command and control of economic activity mixed in with a mostly market economy that uses a price system to allocate resources.

At the end of the day, all government interventions – both good and bad – are the result of a political process. In democracies, the amount of government intervention is, broadly speaking, a reflection of the will of the people.

Encouraging Technology and Innovation

One of the most important jobs of government is helping to promote the invention of new technologies so that we can enjoy higher living standards.

Technology is, in many ways, like any other good that a market can provide. If a profit incentive exists to inventing a new technology, businesspeople are going to figure out a way to invent it, just as they figure out ways to deliver all the other things that people are willing to pay for.

Businesses and governments spend hundreds of billions of research and development funds each year attempting to invent new technologies. Governments provide a good deal of direct support through research grants and university subsidies. But a crucial thing to understand about innovation is the indirect role that governments play not by subsidising new technology but by guarding it.

Catching up quickly

People in western Europe, the United States and Japan are richer than those living anywhere else. The interesting thing about this reality, however, is how long it took for these countries to get so rich.

Because these countries have been at the cutting edge of technology for a long time, the only way they've been able to push out their PPFs and produce more from the same resources has been to invent new technologies. Historically, this adjustment means that living standards in rich countries grow only about 2 per cent per year because they need to invent new technologies in order to raise living standards. At this rate, standards of living take about 30 years to double.

An important thing to realise is that these countries are so much richer than other places not

because of some sudden stroke of luck, but because of a long history of slow but steady progress. That slowness, however, also means that other countries that aren't yet as rich can grow very quickly and catch up to the living standards of the richest nations.

Developing nations like China and India can grow much more quickly because they can jump from using older, less productive technologies to the most productive, cutting-edge technologies. Consequently, they're showing growth rates of 6 to 8 per cent per year. At these rates, living standards double in less than a decade, and, if the trend continues, in only a couple of generations China and India are going to have living standards comparable to those in western Europe, the United States and Japan.

In particular, the patents granted by governments provide a huge economic incentive for both individuals and businesses to innovate. A patent guarantees inventors of new products or business methods the exclusive right to profit from their innovations, usually for about 20 years in most countries.



The fact that economic growth in western Europe and the United States took off 200 years ago, right after patents became widely enforced, is no coincidence. For the first time in world history, a secure financial incentive existed for using your brain to innovate. Before that time, innovating was extremely risky because after all your hard work, others simply copied your invention and sold it without your permission.

Copyrights for literary, musical and cinematic works serve a similar purpose. A great deal more art is produced when artists know that they can make a living off their products. Along these lines, the easy duplication and distribution of digital media on the Internet is a troubling development because it may have weakened artists' ability to charge for the art that they work so hard to produce.

Governments also have a key role to play in encouraging education. Every rich country in the world has a policy of universal primary and secondary education, as well as strong universities. Smart new technologies require smart, well-educated researchers, and you don't get them without good educational systems.

Advanced economies also require smart, well-educated workers to implement the new technologies. Consequently, education must be available to everyone if an economy is to utilise the constant flow of innovative new processes and tools that researchers develop.

Part II

Macroeconomics: The Science of Economic Growth and Stability



'I ignored the rumours about the recession
sweeping the city, and now I am.'

In this part . . .

The chapters in this part introduce you to *macroeconomics*, the study of the economy as a whole, which concentrates on economy-wide factors such as interest rates, inflation and the rate of unemployment. We explain what economists believe causes recessions, and we use the famous Keynesian model to illustrate the policies that economists believe can best be used to fight recessions. Finally, we touch upon the factors that economists believe are essential to promoting sustained economic growth and rising living standards.

Chapter 4

Measuring the Macroeconomy: How Economists Keep Track of *Everything*

In This Chapter

- ▶ Measuring GDP: the total value of goods and services
 - ▶ Deconstructing GDP into $C + I + G + NX$
 - ▶ Understanding why free trade is good for you
-

Macroeconomics studies the economy as a whole. Seen from on high, the production of goods and services is done by businesses or by the government. Businesses produce the bulk of what people consume, but the government provides many goods and services, including public safety, national defence and public goods such as roads and bridges. In addition, the government provides the legal structure within which businesses operate and also intervenes in the economy in order to do things such as regulate pollution, mandate safety equipment and redistribute income from the rich to the poor. (For more on the division of tasks between private businesses and the government, see Chapter 3.)

In order for economists to study the process of production, distribution and consumption with any real understanding, they need to keep track of exactly how much is being produced, as well as where it all ends up. Consequently, economists have developed a huge accounting apparatus to measure economic activity, called the *National Accounts* (internationally known as *National Income and Product Accounts*, or NIPA). This system produces numerous useful statistics, including the famous *gross domestic product* (GDP), which measures the total quantity of goods and services produced in a country in a given period of time.

The system can seem arcane, but knowing how the accounting works is indispensable because it's the basis for all the mathematical models that economists use to understand and predict things such as the business cycle, inflation, economic growth and both monetary and fiscal policy. (We present some of these models in Chapters 6 and 7.) So please make sure that you take, um, proper account of what we're about to show you.

Using GDP to Track the Economy



Gross domestic product, or GDP, is a statistic that calculates the value of all goods and services produced in a given country in a given period of time. In the United Kingdom, the Office of National Statistics calculates and publishes this statistic regularly and revises it just as regularly as more data becomes available, giving us an idea of how much economic activity took place in the previous quarter or year.

GDP is very important because, other things being equal, richer people are happier people (more or less – richer people also moan more loudly about their lot too). We're not saying that money is the only thing that matters in life, but economists evaluate economies by how successfully they maximise happiness, and although money can't buy you love, it can buy you the diamond ring that may attract love, as well as a lot of things that ought to make you happy, such as food, education and holidays (and in economic terms, if these things don't make you happy, you're just better off putting your cash into buying things that do). Consequently, a high and quickly growing GDP is preferable because it reflects lots of economic transactions that provide people with the goods and services they desire. (To examine some reasons why GDP may not *always* reflect increased happiness, see the upcoming section 'The good, the bad and the ugly: All things increase GDP'.)

In Chapters 2 and 3, we discuss how people's fundamental economic goal is to maximise happiness given the limited resources that constrain them. Because people like to consume goods and services, measuring GDP allows economists to quantify, in some sense, how well a country is doing at maximising its citizens' happiness given the country's limited resources. A rising GDP indicates that a country is working out ways to provide more of the goods and services that make people happy.

In this section, we show you how and why the economists who tabulate the National Accounts break up GDP into its constituent parts. Breaking up GDP allows you to analyse each part separately and get a good grasp of the major factors that influence the production of goods and services. But first, we give you a short explanation of what GDP *doesn't* take into account.

Leaving some things out of GDP

The GDP statistic counts only transactions that involve money, so if you look after your elderly parents or if a mother stays home to take care of an infant, that economic activity – though very productive and socially beneficial – doesn't get counted in GDP.

In developed economies, GDP is very good at capturing nearly all the output produced, because almost everything that's produced in these economies is subsequently sold. But in a largely rural and agrarian society of small farmers, most production is for consumption within the household, meaning that the output never makes it to the official GDP statistics that the country's economists keep. As countries undertake the transition from rural agrarian economic structures with lots of household production to market economies where nearly everything produced is sold for money, the GDP appears to rise because a lot of output is being counted for the first time. However, this apparent change may not be an actual increase in output. These limitations can make comparing the GDPs of various countries misleading.

Getting in the flow: Tallying up what counts in GDP

Counting sales where money changes hands can get a little tricky because both a buyer and a seller are involved in every such transaction. The money that the buyer spends has to equal the money that the seller receives. Translated into economist lingo, income has to equal expenditure.

Consequently, you can measure GDP by totalling up all the expenditures in the economy or by counting up all the incomes in the economy. If your calculations are correct, both methods give you the same value for GDP.



When thinking about GDP, you also have to consider the goods and services that are being traded for money. Economists simplify life by saying that all the resources or factors of production of a society – land, labour and capital (see Chapter 3) – are owned by households. *Households* can be made up of one person or several – think in terms of individuals or families. *Firms* buy or rent the factors of production from the households and use them to produce goods and services, which are then sold back to the households. This process sets up a *circular flow* for resources moving from households to firms, and goods and services moving back the other way, as Figure 4-1 shows.

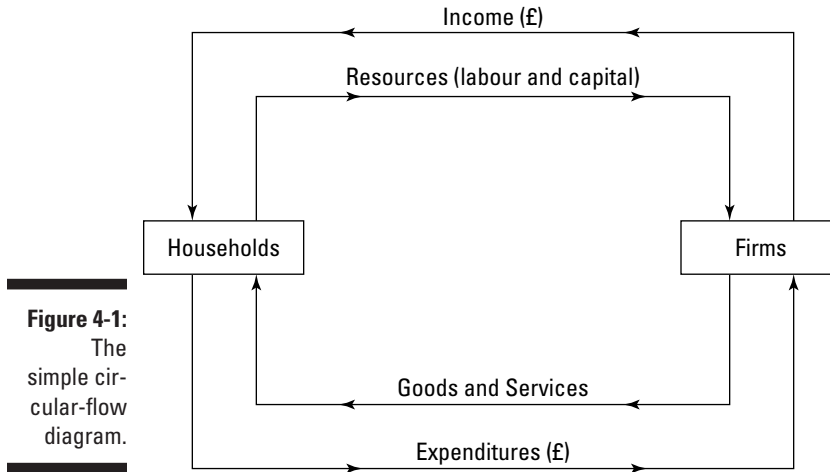


Figure 4-1:
The
simple cir-
cular-flow
diagram.

Moving opposite to the flow of resources and goods are payments in pounds. When the firms buy factors of production from households, they have to pay money to the households. That money is income to the households. And when households buy goods and services from the firms, they pay for those goods and services with money, which shows up in Figure 4-1 as expenditures.

The Science Museum in London has a quite brilliant model of how this works. When we say model, we mean a physical model entirely constructed from clear Perspex pipes and tanks, with coloured water representing the elements and flows. The water flows through the pipes in exactly the same way as money flows around the economy in the model in Figure 4-1.

A key point to understand is that firms are owned by households – and not just billionaires’ households. Firms don’t exist on their own. Ultimately firms are owned by households, either directly, in the case of smaller businesses, or via investment funding from banks or pension funds. In turn, any money that a firm receives when it sells a good or a service flows on as income to some individual or group of individuals. Because of this flow, incomes in Figure 4-1 have to equal expenditures.

Considering flows of income and assets

Although you can use incomes or expenditures to measure GDP, economists prefer to use incomes because governments make both individuals and businesses keep track of every last penny of income they receive so that it can be taxed. This government requirement provides extensive, accurate data about incomes. (We’re sure that you’re not going to have any trouble accepting the truth of this statement, in theory at least.)

Tracing the flow of income

All the income in the economy flows into one of four categories:

- ✓ Labour receives wages
- ✓ Land receives rent
- ✓ Capital receives interest
- ✓ Entrepreneurship receives profits

You may recognise the first three of these categories as being the three traditional factors of production that we list in Chapter 3. Obviously, because you need land, labour and capital to make things, you have to pay for them. That's why some of the income in the economy flows their way. But in a dynamic, competitive economy, you also need people with a willingness to take on business risk and invest in risky new technologies. In order to get them to do so, you have to pay them, which is why some income must also flow to risk-taking entrepreneurs in the form of profits. Hence many economists like to think of entrepreneurship as a fourth factor of production – a factor that must be paid if you want to get stuff produced in a market economy – and this idea has gained wide political currency.

Each of the four payments is a flow of money that compensates for a flow of services needed in production:

- ✓ Workers charge wages for the labour services that they provide.
- ✓ Owners of buildings and land charge rents to tenants for the services that real estate and physical structures provide.
- ✓ Firms wanting to obtain the services of capital, such as machines and computers, must pay for them. This payment is considered interest because, for example, the cost of obtaining the services of a £1,000 piece of capital equipment is the interest payments that a firm must make on a £1,000 loan to buy that piece of equipment.
- ✓ And, finally, the firm's profits must flow to the entrepreneurs and owners of the firm, who take on the risk that the firm may do badly or even go bankrupt.

If all this appears unrealistic, just remember that the purpose of this model is to help us understand the complexity of an economy rather than provide an exact depiction.

Taking assets into consideration

What happens to the flow of income if a firm buys its land and office space rather than rents it? Or if a firm owns its capital outright rather than borrows

money to buy it? If a firm owns these things, it no longer has to pay a flow of money in order to obtain a flow of services. Do expenditures still equal incomes?

No need to fear: incomes still equal expenditures. However, you have to do some fancy accounting to see how this is the case. The key to this balancing act is understanding what constitutes an asset.



An *asset* is something durable that isn't directly consumed but that gives off a flow of services that you do consume. For example, a house is an asset because it provides shelter services. You don't consume the house; you consume the services it provides. Similarly, a car is an asset because, although you don't consume the car itself, it provides transportation services.

You often have a choice between buying an asset outright and thereby owning all the future services that the asset provides, or letting someone else own the asset and sell you the services as they're produced. For example, you can buy a house and thereby get all future shelter services that the house provides, or you can rent the house and get those same services by paying for them each month. For this reason, an asset is considered to be a *stock*, while the services it provides are referred to as a *flow*.

For all assets that a firm owns, accountants put a money value on the services that the assets provide based on what those same services would have cost if the firm had rented them. They can then divvy up the firm's total income, calling some of it rent, some of it interest and some of it profits, as though the owners of the firm are getting three streams of income.

Because the firm's owners provide the money to buy the firm's assets, part of their income is compensation for providing these goods and services, and the rest of their income is counted as compensation for providing entrepreneurship and taking on risk. Consequently, all the money expended on goods and services flows as income to somebody for providing land, labour, capital or entrepreneurship (the four friendly factors of production). This methodology allows economists to keep saying that incomes equal expenditures even if firms own their own assets.

Following the funds, around and around

The simple circular-flow diagram of Figure 4-1 captures the fact that an income exists for every expenditure. However, because the diagram divides the economy only into firms and households, it misses a lot of the action that goes on in the real world. In Figure 4-2, you can see a much more realistic and detailed circular-flow diagram that divides the economy into firms,

households and the government, with these entities making transactions through the following three markets:

- **Markets for factors of production** are where money is exchanged to purchase or rent the land, labour, capital and entrepreneurship used in production.
- **Financial markets** are where people who want to lend money (savers) interact with those who want to borrow money (borrowers). In this market, the supply and demand for loans determine the *interest rate*, which is the price you have to pay to get someone to lend you their money for a while. Because most governments run deficits (in other words, they're always in the hole) and have to borrow a lot of money, they're major players in the financial markets.
- **Markets for goods and services** are where people and the government buy the stuff that firms make.

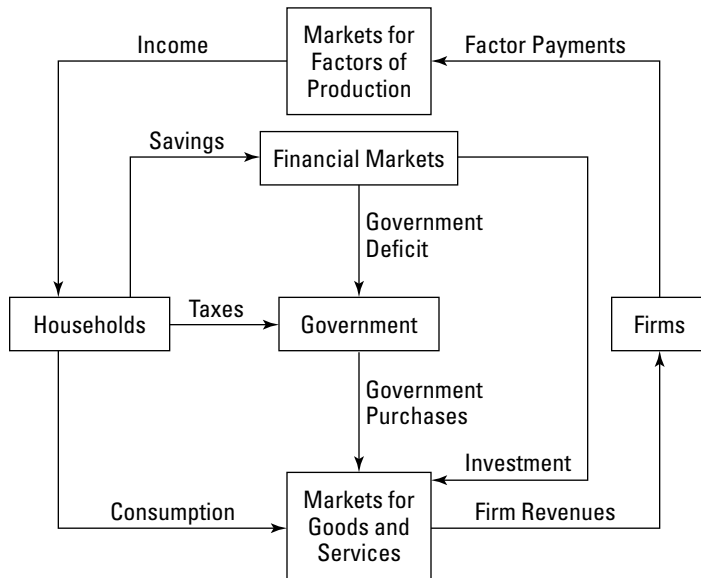


Figure 4-2:
The detailed circular-flow diagram.

In Figure 4-2, arrows show the flows of money throughout the economy. Firms make payments – rent, wages, interest and profits – to households to obtain the factors of production – land, labour, capital and entrepreneurship. (See the previous section ‘Considering flows of income and assets’ for more information on these factors of production.) Households take the income they get from selling these factors and use it to pay for goods and services,

to pay taxes or to save. The government buys goods and services using the tax revenues it takes in or the money it borrows in the financial markets. The financial markets also provide money for corporations to make investments. This money adds to what firms get from selling goods and services to households and the government.

Note that not all transactions in the financial markets are relevant to the calculation of GDP. GDP measures currently produced output, and most transactions in the financial markets are trading property rights for stuff produced long ago. (As an analogy, a house that was built 30 years ago has nothing to do with current production, so the sale of the house doesn't factor into this year's GDP. Only the sales of newly constructed houses figure into this year's GDP.)

Counting stuff when it's made, not when it's sold

Newly produced output is counted as part of GDP as soon as it's produced, even before the output gets sold. That makes keeping track of the money associated with new production a little tricky.



For example, as soon as construction on a new house is completed, its market value of £300,000 (or £1 million if you live in London) is estimated and counted as part of GDP right then, even though the house may not be sold for months. Suppose construction was completed on 29 December 2008, adding £300,000 to 2008's GDP. If the house is subsequently sold on 21 February 2009, it doesn't count in 2009's GDP because double counting isn't allowed.

When sold, the house is considered old property and not new production. Economists just say that the property right to this (now old) house has changed hands from the builder to the new owner. Because trading old assets obviously involves no new production, it doesn't count in GDP.

This accounting convention applies to firms producing any sort of output good whatsoever. If Sony produces a TV on 31 December 2008, the value of that TV is counted in 2008's GDP, even though it won't be sold to a customer until the following year. A handy way to think about this is to imagine that Sony builds the TV and then, in effect, sells it to itself when it puts the TV into inventory. This 'sale' is what is counted in GDP for the year 2008. When the TV is later sold from inventory to a customer, the process is just an exchange of assets (trading the TV for cash).

The fact that output is counted when produced, rather than when sold, is a red flag when interpreting GDP statistics to gauge the health of the economy.

High GDP means only that a lot of stuff is being produced and put into inventory, and not necessarily that firms are selling lots of stuff. (This fact is actually part of a more general problem in interpreting accounts and can apply to firms as well as the economy as a whole.)

In fact, GDP can be high and the economy can be about to go into a recession, because inventories are piling up and managers are soon going to cut back on production in order to get inventories back down to target levels. Consequently, economists who try to forecast where the economy is heading pay much more attention to inventory levels than they do to last quarter's GDP.

The good, the bad and the ugly: All things increase GDP

Generally speaking, higher GDP is better than lower GDP because more output produced means higher potential living standards, including better healthcare for the sick and more money to aid the needy.

But higher GDP doesn't guarantee that happiness is increasing because GDP often goes up when bad things happen. For example, if a major flood destroys a large section of a city, GDP goes up as reconstruction kicks into gear and lots of new output is produced to replace what was destroyed. But, of course, everyone would prefer that the flood not happen in the first place.

Similarly, higher GDP may be possible in certain situations only if you're willing to tolerate more pollution or greater income inequality. Countries experiencing rapid economic development and quickly rising living standards often also get dirtier environments, as well as more social unrest because some people are getting richer much faster than others. The GDP number doesn't reflect these negative conditions.

GDP also doesn't count the value of leisure. Many of your favourite times have probably come about when you weren't producing or consuming anything that would count in GDP – sitting on the beach, climbing a mountain, taking a walk. Moreover, an increase in GDP often comes at the price of sacrificing these leisure activities – meaning that when you see an increase in GDP, overall well-being or happiness hasn't necessarily improved. A reasonable question to ask is whether relatively 'disappointing' European growth rates are actually down to the good burghers of the eurozone placing a higher valuation on leisure. Research also suggests that citizens of the eurozone are relatively happier than their non-eurozone British friends. These two points may not be coincidental!

So although policies that raise GDP are generally beneficial for society, the costs involved in creating the rising output must always be examined.

Introducing the GDP Equation

So far in this chapter, we've only *introduced* you to GDP. Now we want you and GDP to make friends so that you can understand all GDP's little secrets – in particular, its constituent parts and how they behave. Although interesting in its own right, this discussion is doubly useful because it makes the standard Keynesian macroeconomic model (which we introduce in Chapter 6) much easier to understand and manipulate.

The Keynesian model was first developed in 1936 by Cambridge University economist John Maynard Keynes in his book *The General Theory of Employment, Interest and Money*. His text was hugely influential – so influential, in fact, that it led to macroeconomics becoming a separate field of study for economists.



Keynes's book was a response to the Great Depression of the 1930s. Because he felt that government policies designed to fight that economic downturn should focus on getting people to increase their expenditures on goods and services, Keynes began his model by using an equation that measures GDP by adding up expenditures.

In the section 'Getting in the flow: Tallying up what counts in GDP', earlier in this chapter, we explain that you can measure GDP by adding up all the expenditures made on purchasing goods and services or by adding up all the incomes that are derived from producing goods and services. The two numbers have to be equal. So this switch to counting up GDP using expenditure is totally legitimate. (This expenditure method is also the perfect opportunity for you to understand the economy from the point of view of where money gets spent, as opposed to who gets to keep what's earned.)

The expenditure equation for totalling up GDP adds together the four traditional expenditure categories – consumption (C), investment (I), government (G) and net exports (NX) – to equal the value in pounds (or whatever currency a given country is using) of all goods and services produced domestically in that period, or the GDP (Y). In terms of algebra, the equation looks like this:

$$Y = C + I + G + NX \quad (1)$$

Although the following sections go into more detail, here's a quick look at the four expenditure variables that total up to GDP:

- ✔ **C stands for consumption expenditures** made by households on goods and services, whether domestically produced or produced abroad.
- ✔ **I stands for investment expenditures** made by firms on new capital goods including buildings, factories and equipment. *I* also contains changes in inventories, because any goods produced but not sold during a period have to go into firms' inventories and are counted as inventory investments.
- ✔ **G stands for government purchases of goods and services** (they've got to buy paperclips).
- ✔ ***NX* stands for net exports**, which is defined as all a country's exports (*EX*) minus all its imports (*IM*), or $NX = EX - IM$. *EX* is the value in pounds of our output that foreigners are buying. *IM* is the value in pounds of their output that we're buying.

These four expenditures give us GDP because, as a group, they buy up every last bit of output produced in our country in a given period.

C is for consumption

Household consumption spending accounts for about 67 per cent of GDP – far more than the other three components combined. Many factors affect how much of their income households decide to spend on consumption and how much of it they decide to save for the future.

Microeconomists spend a lot of time studying the various factors that affect such decisions, including expectations about whether the future looks bright or dark and how high or low the rates of return are on savings. (See Part III of this book for everything you ever wanted to know about microeconomics.) Macroeconomists, on the other hand, step back from these factors because, when studying the economy as a whole, what matters is how much total consumption exists rather than why households happen to choose that particular level.

Macroeconomists model consumption very simply, as a function of people's after-tax, or disposable, incomes. You can derive disposable income algebraically using this handy three-step process:

1. **Start with Y , the total income in the economy.** In Keynes's equation, Y equals total expenditures, but because income equals expenditures, you can use it for income as well. Remember that any money expended by you is income to someone else.
2. **Figure out how much tax people have to pay.** For simplicity, assume that the only tax is an income tax and that the income tax rate is given by t . For example, $t = 0.25$ means a tax rate of 25 per cent of people's incomes. Consequently, the total taxes that people pay, T , is given by $T = t \times Y$.
3. **Subtract people's taxes, T , from their incomes, Y , to figure out their after-tax incomes.** Economists refer to this amount as *disposable income* and write it algebraically as Y_D . Subtracting taxes from income looks like this:

$$Y_D = Y - T = Y - t \times Y = (1 - t) \times Y \quad (2)$$

After you derive disposable income, you use a very simple model to figure consumption expenditures made by households. The model says that consumption, C , is a function of disposable income and a couple of other variables, C_o and c .

$$C = C_o + c \times Y_D \quad (3)$$

Lowercase c is called the *marginal propensity to consume*, or MPC, where c is always a number between 0 and 1 that indicates the rate at which you choose to consume income rather than save it. For example, if $c = 0.9$, you consume 90 pence of every £1 of disposable income that you have after paying taxes. (You save the other 10 pence.)

The actual value of the marginal propensity to consume, c , is determined by the individual and varies from person to person depending on how much of their disposable incomes they like to save. But what is C_o ? Think of it as how much people consume even if they have zero disposable income this year. (If you assume that $Y_D = 0$ in the equation $C = C_o + c \times Y_D$, that equation reduces to $C = C_o$.) But where does the money come from to pay for C_o if you have zero disposable income? It comes from your personal savings, which you've piled up over the years. Economists call this dis-saving.

The overall equation $C = C_o + c \times Y_D$ says that your total consumption expenditure in an economy is your emergency level (when you have zero income) C_o plus a part of your disposable income given by $c \times Y_D$.

For the rest of this book, we assume that the equation $C = C_o + c \times Y_D$ is a good enough model of how consumption expenditures are determined in

the economy. The equation isn't perfectly realistic, but it does show that consumption is reduced by higher tax rates and that people make a decision about how much of their disposable incomes to save or consume. (And in any case, when you understand the model, you can play about with any number of refinements, which is often how economists get into publications!) The equation allows us to analyse the effects of policies that change tax rates and the effects of other policies that encourage people to spend higher or lower fractions of their incomes.

I is for investment in capital stock

Investment is vitally important because the economy's capacity to produce depends on how much capital is available to make output. The capital stock increases when firms purchase new tools, buildings, machines, computers and so on to help produce consumption goods. Investment is a flow that increases the capital stock of the economy.

But, of course, capital wears out as it's used, by rusting, breaking down or being thrown away when it becomes obsolete. Economists call all these flows that decrease the capital stock *depreciation*.

Naturally, firms must make some investments just to replace the capital that has depreciated. But any investment in excess of depreciation causes the overall size of the capital stock to increase, creating more potential output for people to consume.



The flow of investment spending over any period of time depends on the comparisons that firms make between the potential benefits and the costs of buying pieces of capital. The potential benefits are measured in terms of potential profits, and the costs of buying are measured by the interest rate, regardless of whether or not a firm takes out a loan to buy a given piece of capital.

Why does the interest rate matter so much? Naturally, if a firm needs to take out a loan to buy capital, higher interest rates mean that the firm is less likely to borrow money because of the high loan-repayment costs. However, even if a firm has enough cash on hand to buy a given piece of equipment, higher interest rates force the firm to decide between using the cash to buy the equipment and loaning it out to someone else. The higher the interest rates, the more attractive loaning the money becomes. Consequently, higher interest rates discourage investment regardless of whether firms have to borrow to fund investment. (See Chapter 2 for the reasons why we look at the opportunity cost of investing rather than just the interest rate cost.)

Economists model the amount of investment expenditure that firms desire to make, I , as a function of the interest rate, r , which is given as a percentage. The equation that we use here is standard in introductory books on macroeconomics (although notation does vary from book to book):

$$I = I_0 - I_r \times r \quad (4)$$

This equation is similar in spirit to the consumption equation in the previous section except for the minus sign, which indicates that when the interest rate rises, I falls.

The parameter I_r tells you how much I falls in an entire economy for any given increase in interest rates. For example, suppose that r rises by one percentage point. If I_r is, say, 10 billion, you know that each one percentage point increase in interest rates decreases investment by £10 billion.

The parameter I_0 tells you how much investment occurs if interest rates are zero. In truth, interest rates very rarely fall all the way to zero, but suppose that they do. Then the second term in the equation equals zero, leaving you with $I = I_0$.

The equation as a whole says that if interest rates were zero, investment expenditures max out at I_0 . But as interest rates rise above zero and keep on rising, investment falls more and more. In fact, rates potentially rise so high that investment spending falls to zero. (In the 1990s, the Japanese government used zero interest rates as part of its policy to deal with a protracted slump in the Japanese economy.)

The relationship between rates and investment is one reason why the government's ability to set interest rates has great bearing on the economy. By setting interest rates, the government can determine how much businesses want to spend buying investment goods. In particular, if the economy is in a recession, the government can lower interest rates in order to raise firms' expenditures on investment and (we hope) help improve the economy.

G whizz! Government, that is

In most countries, a huge portion of GDP is consumed by government. The size of that portion is a matter of some variation. According to the international Organisation for Economic Co-operation and Development, it stands at 42 per cent for the US and 48 per cent on average for the EU (it's 46 per cent in the UK). If you take away *transfer payments* – what gets paid out of tax revenues to

transfer in, for example, benefits – which is what we do in this model, the figures are 23 per cent for the US and 27 per cent for the UK.



A government gets the money to buy all that output from taxation and borrowing. If a government's tax revenues are exactly equal to its expenditures, it has a *balanced budget*. If tax revenues are greater than expenditures, a government is running a *budget surplus*. But if expenditures exceed tax revenues, which can happen when a government borrows the difference on the financial markets, that government is running a *budget deficit*.

Governments borrow by selling bonds. A typical bond says that in exchange for £10,000 right now, the government promises to give you back £10,000 in ten years and, in the meantime, pay you £1,000 per year for each of the intervening years. If you accept the deal and buy the bond, you're in effect lending the government £10,000 right now and getting a 10 per cent per year return until the government returns your £10,000 in ten years.

A huge amount of political negotiation goes into determining how much a government is going to spend in a given year. Typically, these negotiations take place in the run up to the autumn financial report by the Chancellor of the Exchequer, but discussion goes on pretty much the whole year. Many groups lobby for special programmes to benefit their hometown or their industry, and no matter what, governments have to provide for essential governmental functions like national defence and law enforcement.

However, economists largely ignore the political machinations that go into determining government expenditures because the economic effects of government expenditure, G , depend on how big the expenditure turns out to be and not on how it got to be that size. Consequently, for the rest of this book, we make the simplifying assumption that government expenditures can be denoted as:

$$G = G_o \quad (5)$$

That is, G is equal to some number determined by the political process, G_o . This number may be high or low, depending on politics, but in the end you care only about how big or small it turns out to be and can ignore where it came from.

G includes only government expenditures on newly produced goods and services, and doesn't include government expenditures that merely transfer money from one person to another. For example, when the government taxes you and gives the money to a poor(er!) person, that transaction has nothing

to do with currently produced goods and services and consequently doesn't count as part of G . So, remember that when we talk about G , we're talking about only the government's purchases of currently produced goods and services.

NX : Exports less imports

When your country sells domestically made goods and services to someone or some firm in another country, such sales are called *exports*, or EX . When someone in your country buys something produced abroad, such purchases are called *imports*, or IM . *Net exports*, or NX , is simply the total value of all exports minus the total value of all imports during a given period of time. When using the expenditure method for totalling up GDP, you add in net exports, NX .

But why only *net* exports? Good question, and economists typically do a lousy job explaining why you have to subtract imports from exports in this equation. Here's the reason.

The whole point of totting up expenditures to get GDP is to figure out how many total pounds were expended on products made within your own country's borders. Most of that expenditure is made by locals, but foreigners can also expend money on your products. That's exactly what happens when they pay you for the goods that you export to them. Consequently, you have to add in EX if you want to get a correct measure of expenditures made on stuff you produce domestically.

You have to subtract your imports of foreign goods because you must differentiate the total expenditures that domestic residents make on *all* goods and services from their expenditures on *domestically made* goods and services. Total expenditures on all goods and services, both domestic and foreign, are C (see the earlier section 'C is for consumption'). If you want to get just the part that's spent on domestically made stuff, you have to subtract the value of imports, IM , because all money spent on imports is money that's *not* spent on domestically made goods and services. So $C - IM$ gives the amount of money that domestic residents spend on domestically produced output.

The result is that you can write your GDP expenditures equation that totals up all expenditures made on domestically produced output as follows:

$$Y = C - IM + I + G + EX \quad (6)$$

But the equation is normally rearranged to put the exports and imports next to each other like this:

$$Y = C + I + G + EX - IM \quad (7)$$

The reason for rearranging is because $EX - IM$ quickly reveals your country's *trade balance*. When $EX - IM$ is positive, you're exporting more than you're importing; when it's negative, you're importing more than you're exporting. Economists like it when the maths is presented in a way that tells a little story.

International trade is hugely important, and you need a good understanding of not only why a trade balance can be positive or negative, but also why you shouldn't necessarily worry if that balance is negative rather than positive. We cover this topic in the next section.

Understanding How International Trade Affects the Economy

Modern countries do a huge amount of trading with other countries – so huge, in fact, that for many countries imports and exports are equal to more than 50 per cent of their GDPs. So, now is a good time to focus a little more deeply on the NX part of the GDP expenditure equation, $Y = C + I + G + NX$.

Understanding how international trade affects the economy is absolutely essential if you hope to have a complete understanding of macroeconomics. This understanding is also important because politicians are constantly suggesting policies such as tariffs and exchange-rate controls that are aimed squarely at international trade – the effects of which reverberate throughout the domestic economy.

This section explains why trade deficits (negative values of NX) aren't necessarily bad and just why engaging in international trade – even when it means sustaining trade deficits – is typically hugely beneficial.

Trade deficits can be good for you!

If your exports exceed your imports, you have a *trade surplus*, whereas if your imports exceed your exports, you have a *trade deficit*. Unfortunately, the words *surplus* and *deficit* carry strong connotations that make it sound like surpluses are necessarily better than deficits, which is simply untrue –

not that you'd know it from the rhetoric that politicians throw around. They make it sound as if trade deficits are always bad and always lead to calamity.



To understand why the politicians are wrong (as if you needed any convincing), consider an example of two individuals who want to trade. Each person starts with £100, and each produces a product for sale. The first person grows and sells apples for £1 each. The second person grows and sells oranges, also for £1 each. Each of them produces 50 pieces of fruit.

Next, suppose that the apple grower really likes oranges and wants to buy 30 of them for £30, and that the orange grower wants to buy 20 apples for £20. Each person is happy to satisfy the other's desires, so the apple grower spends £30 buying oranges from the orange grower, and the orange grower spends £20 buying apples from the apple grower.

These trades shouldn't cause any alarm bells to ring. When people start looking at the trades using the terms *trade surplus* and *trade deficit*, however, they often come to the false conclusion that only one of the growers benefits from the trades when, in reality, they were both quite eager to trade in this way.

To see where the confusion arises, notice that in the vocabulary of international trade, the apple grower exports only £20 worth of apples but imports £30 worth of oranges. At the same time, the orange grower exports £30 worth of oranges but imports only £20 worth of apples. As a result, you have a situation in which the apple grower is running a £10 trade deficit and the orange grower is running a £10 trade surplus.

Does this mean that the apple grower is worse off than the orange grower? No. Each person started with £150 worth of stuff: their respective £100 cash piles plus £50 each worth of fruit. When they finish trading, they each still have £150 worth of stuff. The apple grower has £90 of cash plus £30 worth of apples and £30 worth of oranges. The orange grower has £110 of cash plus £20 worth of oranges and £20 worth of apples.

Saying that their trading has made one of them poorer is way off the mark. In fact, both of them are happier with their arrangements of wealth after trading than they were before trading because their trades were voluntary. If the apple grower would have been happier keeping his initial holdings of £100 cash and 50 apples, he wouldn't have traded for oranges. And the same applies with the orange grower.



As long as international trade is voluntary, all trades enhance happiness (that is, create gains, although those gains may not be equally divided). To concentrate on whether a trade deficit or surplus exists is to completely miss the point that international trade is simply a rearrangement of assets between countries that makes everyone happier. Even the country running the trade deficit is happier.

Considering assets – not just cash

To people who hate trade deficits, the fact that the apple grower's cash pile falls from £100 before the trade to only £90 after the trade looks spooky, because these people focus totally on the fact that the apple grower is £10 poorer in terms of cash after the trading. And they're even more peeved because that £10 ends up with the orange grower, giving her a commanding £110 to £90 advantage in terms of cash piles.

This perspective misses the fact that the apple grower's overall wealth is still £150 and that he now has a distribution of assets that is more pleasing to him. But, if you point this out, deficit haters respond by asking you what happens after the apple grower eats his 30 apples and 30 oranges and after the orange grower eats her 20 apples and 20 oranges. In the end, all that the fruit growers have left are their respective cash piles. Because the apple grower has £20 less cash than the orange grower, he must be worse off by running a trade deficit.

Again, this reasoning misses the point that the apple grower was happier trading and ending up with £90 of cash than he would have been not trading and ending up with £100 in cash. If it weren't for trade, he'd have had a very boring diet of only apples.

Opponents of trade deficits really make things seem scary when they start talking about land trading hands due to international trade. ('Oh no, the foreigners are taking over the country!') To see what they mean, imagine that instead of starting with £100 each, the fruit growers each start with 100 hectares of land worth £1 per hectare. The only way for the apple grower to come up with £10 of cash to pay for his trade deficit is by selling 10 hectares of land to the orange grower. That is, the overall exchange that they engage in is 20 apples plus 10 hectares of land worth a combined £30 in exchange for 30 oranges worth £30. Because 10 of the apple grower's hectares of land now belong to the orange grower, deficit haters think the apple grower sold out his country – literally.



Such transfers of property do happen in real life. During the 1980s, the United States ran huge trade deficits with Japan. The result was that Japanese corporations and individuals ended up owning many famous US buildings and companies. This situation really spooked many jingoistic US politicians (and Harvey Keitel's detective character in the 1993 film *Rising Sun*), but they missed the point that all trading in life – be it with foreigners or fellow citizens – results in increased overall welfare. After all, what good is keeping all your 100 hectares of land if you're happier trading 10 of them for foreign-made goods? Or, in the case of the United States during the 1980s, what good is continuing to own Times Square or Columbia Pictures if you'd rather trade them

for Honda Accords and Sony VCRs? (The anti-Japanese hysteria at the time was even sillier given that the largest group of foreign owners of US property was, and still is, the British!) This example is not an isolated one. Politicians in all countries are happy to embrace this line of reasoning, for both noble and utterly ignoble reasons. Some of them even do it seriously.

Policy debate in the UK largely accepts that foreign investment is, on the whole, beneficial to the national economy, although people raise concerns when foreign market protections prevent UK businesses from buying foreign companies. In 2006, a record number of foreign companies bought UK companies. Proposed and completed purchases included Dubai Ports taking over P&O, Spanish company Ferrovial buying BAA and an Icelandic consortium purchasing West Ham FC. The principle that we should accept these investments as cash inflows to fund more productive activity has hardly been challenged, although each purchase did generate some discussion.

Much to the chagrin of economists, the argument that the point of trade is to make one *happier* doesn't always fly well. A lot of people view trade as an antagonistic contest to dominate other countries by constantly running trade surpluses so that they eventually own all the other person's assets. To this end, they argue for restrictions on trade designed to rig trade relations so that their own countries *always* run surpluses.

The problem here is that if your country puts barriers in place, so can others. That reduces the total volume of trade and leaves everyone less happy. At the extreme, tit for tat restrictions lead to trade wars, which are never in anyone's interest.

Consequently, for the last 50 years, national governments have increasingly pushed for fewer and fewer restrictions on international trade (via forums such as the General Agreement on Tariffs and Trade, or GATT, and the World Trade Organisation, or WTO). This *free trade* movement has resulted in hundreds of millions of new jobs and a vast improvement in living standards and happiness because people all over the world are free to trade and buy whatever they want to make them most happy – even if that means buying from a foreigner.

Does this situation mean that we shouldn't take seriously the arguments of anti-globalisation protesters? No, because as we note earlier, although trade makes both parties happier, one party may be made much happier than the other, and that may be because the trade takes place on unfair terms. People have a number of arguments with the system of international trade, but very few of these arguments reduce to saying that trade as a whole is bad. Instead, the disagreements concern the way in which trade is conducted and the systems that are in place for the governance of trade.

Wielding a comparative advantage

The argument that even countries running trade deficits are better off because they get to consume a mix of goods and services they wouldn't get otherwise rests solely on the benefits of trading things that have already been produced. But an even better argument for international trade is the fact that it actually increases the total amount of output produced in the world, meaning that more output per person results, and overall living standards rise.

This argument, known as *comparative advantage*, was developed by the English economist David Ricardo in 1817 as a forceful rebuttal against import tariffs known as the *Corn Laws*, which heavily taxed imports of foreign-grown grain at the time. These laws kept the price of grain high, and so the nobility that owned the vast majority of farmland favoured retaining them. Naturally, the poor were opposed because the laws drove up the price of their basic food supply: bread.

Ricardo pointed out that abolishing restrictions on international trade would, in addition to helping England's poor, actually make England and all the countries it traded with richer by encouraging them to specialise in the production of goods and services that each of them was able to produce at the lowest possible cost. He demonstrated that this process of specialisation would increase total worldwide output and thereby raise living standards.



The logic behind the comparative advantage argument is most easily understood by thinking in terms of people instead of countries. Consider a patent lawyer named Heather and her brother Adam, who works as a bike mechanic. Heather is very good at filing patents for new discoveries, but she's also very good at repairing bicycles. In fact, she's faster at repairing them than her brother. On the other hand, Adam can file patents as well, though not as quickly as Heather. Table 4-1 lists how many bike repairs and patent filings each of them can do in one day if they put all their efforts into only one of the activities.

<i>Person</i>	<i>Patent Productivity</i>	<i>Bike Repair Productivity</i>
Heather	6	12
Adam	2	10

In one day's work, Heather can produce 6 patents or repair 12 bikes, whereas Adam can file 2 patents or repair 10 bikes. Heather is more efficient than her brother at producing both patents and bike repairs because she can convert one day's labour into more of either good than Adam.

Economists say that Heather has an *absolute advantage* over Adam at producing both goods, meaning that she's the more efficient producer of both; with the same amount of labour input (one workday), she can produce more than her brother. Before David Ricardo came along and explained comparative advantage, the only thing anyone knew to look at was absolute advantage. And when they saw situations like that of Heather and Adam, they concluded (incorrectly) that because Heather is more efficient than Adam at both tasks, she doesn't need to trade with him.

In other words, people used to believe, incorrectly, that because Heather is better than Adam at repairing bikes, she should not only work hard as a patent attorney filing lots of patents, but also fix her own bike whenever it breaks down. Ricardo pointed out that this argument based on absolute advantage is bogus and that Heather should, in fact, *never* fix bikes despite the fact that she's the most efficient bike repairer around. The nifty thing that Ricardo realised is that the world is better off if each person (and country) specialises.



The key insight of comparative advantage is that the proper measure of cost when considering whether Heather should produce one good or the other isn't how many hours of labour input it takes her to produce one patent or one bike repair (which is the logic behind absolute advantage). Instead, the true cost is how much production of one good you have to give up to produce a unit of the other good.

To produce one patent, Heather must give up the chance to repair two bikes. In contrast, to make one patent, Adam has to give up the chance to repair five bikes. So, Heather is the lower-cost producer of patents and, therefore, should specialise in filing patents. And Adam should specialise in bike repairs because he's the lower-cost producer of bike repairs.

On a larger scale, countries should specialise in the production of goods and services that they can deliver at lower costs than other countries. If countries are free to do this, everything that's produced comes from the lowest-cost producer. Because this arrangement leads to the most efficient possible production, total output increases, thereby raising living standards.

Politicians often argue that countries shouldn't be dependent on other countries for various goods and services. Any policy that takes this warning seriously by impeding trade and specialisation increases costs and makes total output fall.

By letting comparative advantage guide who makes what, free trade increases total world output and thereby raises living standards. Under free trade, each country specialises in its area(s) of comparative advantage and then trades with other countries to obtain the goods and services it desires to consume.

Don't be tricked by absolute advantage. As you can see in this section's example, Heather has an absolute advantage at everything but has a comparative advantage only at filing patents. Having an absolute advantage means that you can make something at a lower cost as measured in inputs. (For example, Heather requires fewer hours of labour input to file a patent than Adam.) However, what matters in life isn't inputs but outputs – the things that people actually want to consume. By focusing on costs as measured in terms of alternative types of output that must be given up to produce something, comparative advantage ensures that you're focusing on being efficient in terms of what really matters: output.

Chapter 5

Inflation Frustration: Why More Money Isn't Always a Good Thing

In This Chapter

- ▶ Risking inflation by printing too much money
 - ▶ Measuring inflation with price indexes
 - ▶ Adjusting interest rates to take account of inflation
-

Inflation is the word economists use to describe a situation in which the general level of prices in the economy is rising. This situation doesn't mean that every price of every good is going up – a few prices may even be falling – but the overall trend is upward. Typically, the trend is for prices to go up only a small percentage each year, but people dislike even mild inflation because, face it, no one likes paying higher prices. Mild inflation also causes problems such as making retirement planning difficult. After all, if you don't know how expensive things are going to be when you retire, calculating with any certainty how much money you need to be saving right now is difficult.

Things can go from bad to worse if inflation really gets out of control and prices begin rising 20 or 30 per cent per month – something that has happened in more than a few countries in the last century. Such situations of *hyperinflation* usually accompany a major economic collapse featuring high unemployment and a major decrease in the production of goods and services. (For more about prices and how they affect the economy, see Chapter 6.)

The good news, however, is that economists know exactly what causes inflation and precisely how to stop it. The culprit is a money supply that grows too quickly, and the solution is simply to slow or halt the growth of the money supply. Unfortunately, some political pressure is always exerted in favour of inflation so that simply knowing how to prevent inflation doesn't necessarily mean it isn't going to develop.

In this chapter, we share some things about money and inflation that you may not already know, including why governments are often tempted to print a lot of money to pay for budget deficits, why doing so is actually a form of taxation and why certain groups encourage the government to print a ton of money. We also show you why printing lots of money causes inflation, how to measure inflation and how to measure the effect of inflation on interest rates. The one thing we don't tell you is how to print your own funny money – this book ain't *Counterfeiting For Dummies*.

Buying an Inflation: The Risks of Too Much Money

We can't overstate how important money is to the proper functioning of the economy. Without money, you'd waste most of your time *bartering*, or arranging trades of one good for another – you know, like in primary school ('I'll trade you my apple for your cake!'). Bartering works well only in the rare circumstance that you run into somebody who has what you want and who wants what you have.

Money provides a medium of exchange so that you can still trade for the cake from the kid next to you, even if you don't have an apple. Money can be any good, object or thing, but its defining characteristic is that it's accepted as payment for all other goods and services. (Hence the stuff on a note about promising to pay the bearer on demand.) In today's economy, people pay for things using a wide variety of monies, including government-issued coins and cash, cheques drawn on private bank deposits and electronic payments facilitated by credit cards and debit cards. Because it affects nearly every economic transaction that takes place, money is at the heart of *macroeconomics*, the study of the economy as a whole.

Balancing money supply and demand

As with everything in life, balance is essential. If a government prints too much money, prices go up and you get inflation. If a government prints too little, prices go down and you get deflation. But how much money is the right amount? And why does printing too much or too little cause inflation or deflation?



Beating barter: Show me the money!

Historically, people have used a wide variety of things as money:

- ✔ Seashells were used as money in ancient China, throughout the Pacific and also by Native Americans.
- ✔ Boxes of cigarettes were used as money in prisoner-of-war camps during the Second World War.
- ✔ Various agricultural products, such as barley or cattle, were used as money by many cultures.
- ✔ Huge doughnut-shaped stones were used on the island of Yap in the Pacific.

Eventually, most of the ancient world realised that metal made the best money. Metal doesn't

wear out or shatter like seashells; it doesn't get mouldy like barley; and it can easily be carried around in your pocket, unlike giant doughnut-shaped stones. Shaping metal money into coins, though, was a later innovation. The first metal monies had other shapes, with early Celts preferring ring money; ancient Mesopotamians being fond of long, helical ribbons of metal; and the Chinese using metal monies cast in the shapes of knives and spades.

Regardless of the shape or substance, nearly every society fell upon some good or other to serve as money. If they didn't, they were stuck with barter – a fate everyone wanted to avoid. Cases where the local currency has fallen into disrepute tend to involve some reputable foreign currency.

Basically, the value of money is determined by supply and demand (which we discuss in detail in Chapter 8):

- ✔ The *supply* of money is under government control, and the government can very easily print more money any time it wants to.
- ✔ The *demand* for money derives from its usefulness as a means of paying for things and from the fact that having money means not having to engage in barter.

For any given supply of money, supply and demand interact to set a value for each unit of money. If money is in short supply, each piece of money is very valuable; fewer pieces of money translate into fewer chances to avoid having to engage in barter. But if the government greatly increases the supply of money, each individual unit of money loses value because getting enough money together to avoid barter is easy.



Prices and the value of money are *inversely related*, meaning that when the value of money goes up, prices go down (and vice versa). To see how this relationship works, suppose that money is in short supply and is consequently very valuable. Because money's very valuable, it buys a lot of stuff. For

example, imagine that £10 buys 1,000 grams of coffee (that is, you get 10 grams for 10 pence). But if money's very common, each unit isn't very valuable. In this case, say that £10 buys only 100 grams of coffee (that is, you get only 1 gram for your 10 pence and have to pay £1 to get your 10 grams). Therefore, the greater the supply of money, the higher the prices.

The demand for money tends to grow slowly over time; growing economies produce more stuff, and consumers demand more money with which to buy the available stuff. Depending on how a government reacts to consumer demand for more money, three scenarios are possible:

- ✓ If a government increases the supply of money *at the same rate* as the growing demand for money, prices don't change. In other words, if supply and demand for money grow at equal rates, the relative value of money doesn't change.
- ✓ If the government increases the supply of money *faster* than the demand for money grows, inflation results as money becomes relatively more plentiful and each piece of money becomes relatively less valuable. With each piece of money carrying less value, you need more of it to buy stuff, causing prices to rise.
- ✓ If the government increases the supply of money *slower* than the demand for money grows, deflation results because each piece of money grows relatively more valuable. Buying any given good or service requires less money.



You may be wondering if any way exists to know exactly how much inflation you can expect from printing any given amount of extra money. You're in luck! The *quantity theory of money* states that the overall level of prices in the economy is proportional to the quantity of money circulating in the economy. *Proportional* just means that things go up by equal amounts, so the quantity theory can also be stated this way: if you double the money supply, you double prices.

But *why* would any government want to cause inflation or deflation of any size whatsoever? For the answer to that question, read on!

Giving in to the inflation temptation

Inflation of prices is often explained by governments printing more paper money or producing a large amount of cheap-metal coins, which vastly increases the supply of money and makes each piece of money less precious. As sellers demand higher prices to make up for the fact that each piece of money is worth less, you get inflation.

Croesus and Kublai: The kings of money

King Croesus of Lydia is usually given credit for solving the problem of bogus metal money. In the sixth century BC, Croesus issued the first government-certified coins that guaranteed purity and weight. Lydia was located in what is now western Turkey, and soon all the major trading nations of the Mediterranean were using the new Lydian coins because they were by far the most trustworthy medium of exchange available. The new coinage gave Lydian traders a major advantage, and the kingdom soon became very wealthy, so much so that Croesus was considered the richest man in the world – even richer than King Midas (of *Midas touch* fame), whose gold Croesus minted into coins. (That's where we get the phrase, often applied to owners of Chelsea FC, of being as rich as Croesus.)

But coins are hard to carry around in large amounts, and it was up to the Mongolian emperor Kublai Khan to create the first paper money in the 13th century. This paper money was actually a kind of precious-metal certificate; people holding one of these certificates were able to go to a government vault and redeem it for gold. Consequently, the pieces of paper were as good as gold, but a stack of paper was a whole lot easier to carry than a heavy bag of coins.

Paper money was such a radical innovation that when Marco Polo came back from China and told Europeans about it, they laughed, unable to conceive of anything other than gold or silver coins serving as money. Their incredulity was hard to overcome, and after paper money fell out of favour in China, it was centuries before another government issued any again.

So why in the world do governments ever print too much money? Good question. Historically, governments circulate more money in three circumstances:

- ✔ When governments can't raise enough tax revenue to pay their obligations
- ✔ When governments feel pressure from debtors who want inflation so that they can repay their debts using less valuable money
- ✔ When governments want to try to stimulate the economy during a recession or depression

As you find out more about these three reasons for increasing the money supply, keep in mind what we discuss in the previous section: if the supply of money increases faster than the demand for money, inflation results. Consequently, no matter what reason a government has for increasing the supply of money, it runs the risk of inflation. And that's true both for good reasons, such as wanting to help the economy out of a recession, and for bad reasons, such as helping debtors to repay their loans using less valuable money.



Hyperinflation and Hitler

History's most infamous hyperinflation hit Germany in the 1920s, during the economically incompetent Weimar Republic. Hyperinflation so badly ruined the German economy that Germans later voted Adolf Hitler into power because he promised to fix things.

At the end of the First World War, Germany faced the prospect of paying off massive debts taken on during the conflict in addition to all the ongoing costs of running a government. Most of its debts were in its own currency, the German mark.

Because the German government had the exclusive right to produce German marks, the debt proved an irresistible temptation to begin printing money to pay its bills. Soon, all the new money caused a wild hyperinflation. In fact, the rate of inflation in Weimar Germany in 1922 was well over 100 per cent per month – it reached nearly 6,000 per cent by the end of year!

Then things *really* got out of control. Prices went up 1,300,000,000 times (that's not a

misprint!) in 1923. That year, Germans paid 200,000 marks for a loaf of bread and 2 million marks for a kilo of meat. Prices rose so rapidly that waiters at restaurants had to pencil in new prices on menus several times a day. And if you ate slowly, you were sometimes charged twice what was printed on the menu because prices had gone up so much while you were eating! In some places in Germany, people stopped bothering to take the time to count out money. Instead, they tied paper bills into huge bricks and weighed the bricks of cash. For example, it may have cost two kilos of cash to buy a chicken. Although this catastrophe is one of the most famous, it probably wasn't the first (it may well have happened in Ancient Rome too), and it was certainly not the last case. In Brazil, until surprisingly recently, the main newspapers used to print the rate of indexation, so that people were able to calculate that day's prices given the previous day's prices and the rate of inflation, and the last decade has seen hyperinflations in, for example, Zimbabwe.

Paying bills by printing bills: Heading for hyperinflation

Governments almost always have debts, and printing extra money can be a tempting way to pay them. Quite often, a government wants to spend more money than it's collecting in tax revenue. One solution is to borrow the shortfall, but another is simply to print up new bills to cover the difference.

Until very recently, printing new bills was difficult because most of the world's paper currencies were backed by a valuable metal, such as gold. Under this system, every piece of paper money circulating in the economy was convertible into a specific quantity of gold so that anyone holding cash was able to redeem their cash for gold any time they wanted. For example, you were able to go to the Bank of England with a sum of money and exchange it for a precise amount of gold. This so-called *gold standard* made it difficult for the government to devalue the currency by printing too much

money because it first had to get more gold with which to back the new money. Because purchasing gold is expensive, governments were effectively restrained from increasing their money supplies.

A major break occurred in 1971 when, in order to pay for the escalating costs of the Vietnam War, President Nixon took the United States off the gold standard and put the US on the *fiat system*, in which paper currency isn't backed by anything. People just have to accept the currency as though it has value. In fact, *fiat* is Latin for 'Let it be done'. So when you say *fiat money*, you're basically referring to how a government creates money simply by ordering it into existence. The problem with a fiat money system is that nothing limits the number of little pieces of paper that the government can print up to pay its debts.

The trouble with printing money to pay your debts and obligations is that as soon as the money's out there, people spend it, drive up prices and cause inflation. And if you print more and more money, you end up with people offering shopkeepers and producers more and more money for the same amount of goods. The result is like a giant auction where everybody bidding on items keeps getting more and more money to bid with. The more money you print, the less each individual pound, euro, dollar, doubloon or whatever is worth.

If a government gets into the habit of rapidly printing new money to pay its bills, inflation can soon reach or even surpass 20 or 30 per cent per month, a situation referred to as a *hyperinflation*. Economists hate hyperinflations because they greatly disrupt daily life and ruin the investment climate.



Hyperinflation causes people to waste huge amounts of time trying to avoid the effects of rising prices. During the Weimar hyperinflation in Germany (which we discuss in the sidebar 'Hyperinflation and Hitler'), men working at factories were paid two or even three times a day because money lost its value so quickly. Their wives waited at the factories to take the money immediately to the nearest shops, trying to spend the pay before it lost most of its value. Shopping may be fun, but not when you're desperately racing against outrageously rising prices!

Hyperinflation also destroys the incentive to save because the only sensible thing to do with money during a hyperinflation is to spend it as quickly as you can before it loses even more of its value. Those people whose life savings were in German marks during the Weimar hyperinflation soon found that what they had worked so hard to amass had become worthless. And people thinking about saving for the future were greatly discouraged because they knew that any money they saved would soon lose all value. The discouragement of saving causes major business problems because if people aren't saving, no money is available for businesses to borrow for new investments. And without new investments, the economy can't grow.

Feeling printing press pressures: The politics of inflation

Even if the government isn't trying to use inflation to avoid tax increases, one group in particular always pressures it to circulate more money. You may even be a member of this group – they're called *borrowers*.



To understand the politics of inflation, you need to understand that one of the functions of money is as a *standard of deferred payment*. What does that mean? Imagine that you borrow £1,000 to invest on your farm, promising to pay the bank back £1,200 next year. For the past several years, prices in the economy have been stable, and, in particular, the pigs that you raise have sold for £100 each. Essentially, your loan lets you borrow the equivalent of ten pigs with the promise to pay back twelve pigs next year.

But you have an idea. You lobby your MP to lobby the government to print more money. In a collective rush of blood to the head, the Treasury agrees and instructs the mint to print a load more money. All that new money causes an inflation, after which the price of pigs rises to £200 each. Now you have to sell only six pigs to pay back the £1,200 loan, leaving you with more pigs, you pig!

Lenders, of course, oppose the inflationary desires of borrowers. If you were putting money in the bank, you'd do everything in your power to stop the inflation. If the change goes through, not only are your profits ruined, but also you're an outright loser. In the first year, your loan of £1,000 is the equivalent of ten pigs. But after the inflation, you get paid back the equivalent of only six pigs. You take a 40 per cent loss on the value of your loan. Too much inflation, and a lender ends up being a pig in a poke.

As long as economies use money, lenders and borrowers are always going to be lined up against each other, both trying to sway the government.

Stimulating the economy with inflation

A much more legitimate reason for governments to print more money has the very respectable name of monetary policy. *Monetary policy* refers to the decisions a government makes about increasing or decreasing the money supply in order to stimulate or slow down the economy.

We go into monetary policy in detail in Chapter 7, but the basic idea is that if the economy is in a recession, the government may print up some new money and spend it. All the goods and services it buys with the new money stimulate the economy immediately. In addition, all those businesses that received money from the government can now go out and spend that new money themselves. And whoever receives the money from them also goes out and spends it to buy things. In fact, this pattern can theoretically go on forever and stimulate a heck of a lot of economic activity – enough to lift an economy out of a recession.

Inflation, angry farmers and *The Wizard of Oz*

In the second half of the 19th century, US farmers in the newly opened West found themselves deeply in debt to eastern bankers as a result of the technological revolution then sweeping agriculture. Mechanical harvesters, threshers and other pieces of large and expensive farm equipment greatly increased productivity and output, but the ensuing tremendous increase in supply meant that the prices of agricultural goods plummeted.

Farmers were in a bind because although they were receiving less for their output, they had to keep making large payments on the loans that they had taken out to buy all the expensive new farm equipment. Most farmers settled upon the solution to support political candidates who were in favour of moving the United States from a gold standard to a *bi-metallic*, or gold and silver, standard. Foremost among these candidates was Nebraska senator and two-time Democratic presidential nominee William Jennings Bryant. He argued vigorously for backing US paper money with both silver and gold, because the government could then print more currency than if money were backed only with gold. Although he didn't say so directly, what he wanted was a big inflation.

This political fight pitted western farmers against eastern bankers. The eastern bankers eventually won, and the United States stayed on a gold-only standard. Yet, Americans may well have a great cultural legacy of that political

fight over inflation – although most people don't realise it.

In 1964, a professor named Henry Littlefield speculated that the book *The Wonderful Wizard of Oz* was a political work meant to support the farmers' opposition to the gold standard. Dorothy is a young farm girl from Kansas who represents rural US citizens; the Tin Man represents city workers; the Cowardly Lion is William Jennings Bryant, whom the author thought was not a strong enough leader; and the Scarecrow is the US farmer. The four characters travel towards the East on the yellow brick road – a road made of gold – to see the Wizard of Oz, who represents the evil eastern bankers who manipulate the economy by pulling strings and levers behind a curtain. Oz is simply the abbreviation for *ounce*, as in ounces of gold.

After Dorothy and her companions expose the Wizard and the gold standard as frauds, everything is right in the world. The Scarecrow is intelligent, the Lion gets his courage and the Tin Man never has to worry about rusting (that is, being unemployed) ever again. And in the book, Dorothy returns home thanks to her *silver* slippers. According to Littlefield, the film adaptation used ruby slippers because they looked better on film – a decision that may have led Americans to forget that the story may have been intended as much more than a children's book.

If this result sounds too good to be true, that's because it is. And the reason is inflation. When people start spending all that new money, it drives up prices. Eventually, the only effect of the government's good intentions is that prices rise and no additional goods are sold. For example, if the government doubles the money supply, businesses double the prices they charge because each piece of money is worth half as much as before. Consequently,

the total amount of goods and services sold is the same as before because, although twice as much money is being spent, prices are also twice as high.

The sad upshot is that an increase in the money supply stimulates the economy only when the increase is a surprise.

If the government can print the money and start spending it before people can raise prices, you get an increase in the amount of goods and services sold. Eventually, of course, people figure it out and raise prices, but until they do, the monetary stimulus works.

Unfortunately, continuing to fool people is difficult. You can surprise people once, but the second time is harder and the third time harder still. In fact, if the government keeps trying to surprise people, people begin to anticipate the government and raise prices even before the government prints more money. Consequently, most modern governments have decided against using this sort of monetary stimulus and now strive for zero inflation or very low inflation.

Tallying up the effects of inflation

In recent years in the UK, prices have risen only a small amount each year (about 2.5 per cent annually, although the figure has been much higher in the past). However, even moderate inflation causes problems by cutting into the practical benefits of using money instead of barter. You can get a better sense of this fact by looking at the four functions that economists generally ascribe to money and the ways in which inflation screws up each of them:

- ✔ **Money is a store of value.** If you sell a cow today for one gold coin, you should be able to turn around and trade that gold coin back for a cow tomorrow or next week or next month. When money retains its value, you can hold it instead of holding cows, or property or any other asset.

Inflation weakens the use of money as a store of value because each unit of currency is worth less and less as time passes.

- ✔ **Money is a unit of account.** When money is widely accepted in an economy, it often becomes the unit of account in which people write contracts. People start using phrases like ‘\$50 worth of timber’ rather than ‘50 square metres of timber’.

This practice makes sense if money holds its value over time, but in the presence of inflation, using money as a unit of account creates problems because the value of money declines. For example, if the value of money is falling fast, how much timber, exactly, is ‘\$50 worth of timber’?

✔ **Money is a standard of deferred payment.** If you want a cow, you probably wouldn't borrow a cow with the promise to repay two cows next year. Instead, you'd be much more likely to borrow and repay in terms of money. That is, you'd borrow one gold coin and use it to buy a cow, after promising to pay back two gold coins next year.

The progressive devaluing of money during a period of inflation makes lenders reluctant to use money as a standard of deferred payment. Suppose a friend asks to borrow £100, promising to pay you £120 in a year. That seems like a good deal – after all, the interest rate is 20 per cent. But if prices are rapidly rising and the value of money is falling, how much are you going to be able to buy with that £120 next year?

Inflation makes people reluctant to lend money. They fear that when the loans are repaid, the repayment cash isn't going to have the same purchasing power as the cash that was lent. This uncertainty can have a devastating effect on the development of new businesses, which rely heavily on loans to fund their operations.

✔ **Money is a medium of exchange.** Money is a *medium* (literally meaning 'something in the middle') of trade between buyers and sellers because it can be directly exchanged for anything else, making buying and selling much easier. In a barter economy, an orange farmer who wants to buy beer may have to first trade oranges for apples and then apples for beer because the guy selling the beer wants only apples. Money can eliminate this kind of hassle.

But if inflation is bad enough, money is no longer an effective medium of exchange. During hyperinflations, economies often revert to barter so that buyers and sellers don't have to worry about the falling value of money. For example, in a healthy economy, the orange seller can first sell oranges for cash and then trade the cash for beer. But during a hyperinflation, between the time he sells the oranges for cash and buys the beer, the price of beer may have skyrocketed so high that he can't buy very much beer with the cash. During a hyperinflation, economies have to resort to cumbersome bartering. At the very least, if one currency becomes debased, people often tend to use a trusted foreign currency, usually the US dollar, as the medium of exchange.

Another effect of inflation is that it functions as a giant tax increase. This seems strange because you normally think of governments taxing by taking away chunks of people's money, not by printing more money. But a tax is basically anything that transfers private property to the government. Debasing the currency or printing more money can have this effect.



Suppose that the government wants to buy a £20,000 van for a village. The honest way to go about this is to use £20,000 of tax revenues to buy a van. But a sneakier way is to print £20,000 in new cash to buy the van. By printing and spending the new cash, the government has converted £20,000 of private

property – the van – into public property. So, printing new cash works just like a tax. Because printing new money ends up causing inflation, this type of taxation is often referred to as an *inflation tax*.

Not only is the inflation tax sneaky, but also it unfairly targets the poor because they spend nearly all their income on goods and services, the costs of which go up greatly during an inflation. By contrast, because the rich have the opportunity to save a lot of their incomes instead of spending everything they take in, proportionately they're less affected by an inflation tax. By investing their savings in assets (like property) whose prices go up during inflation, the rich can insulate themselves from a great deal of the harm caused by inflation.

Measuring Inflation: Price Indexes

Inflation can cause lots of problems, so in order for the government to keep inflation under control, it needs a way to measure inflation accurately.

As we explain in the earlier section 'Buying an Inflation: The Risks of Too Much Money', the value of money is determined by the interaction of the *supply* of money with the *demand* for money. The supply of money is under the government's control, but the government can't directly ascertain the demand for money, so it has to look at how supply and demand interact in order to determine how much to increase or decrease the money supply:

- ✓ If an inflation is in effect, the government knows that the supply of money is increasing faster than the demand for money. If it wants to tame the inflation, it needs to reduce the supply of money.
- ✓ If a deflation is in effect, the government knows that the demand for money is increasing faster than the supply of money. If it wants to end the deflation, it needs to increase the supply of money.

Because inflation is a *general* increase in prices, the best way to look for it is to see whether the cost of buying a large collection of many different things changes over time. If, instead, you look at only one or two prices, you may end up confusing a *relative* price change for a *general* price change. (A relative price change is when one price goes up relative to the others, which remain unchanged.)

Economists arbitrarily define some large collection of goods and services and refer to this collection as a *market basket*. They then find out how much money is necessary to buy this basket at various times to measure inflation. In the UK several different measures have been used to try to capture the

effect of inflation. Until recently the headline measure was the RPI-X; that is, the Retail Price Index less mortgage interest (to strip out the effect of interest rates, which otherwise complicate the picture too much). Recently, though, a new measure, called the CPI (Consumer Price Index), does roughly the same thing. You can find the CPI figure at the UK National Statistics website (www.statistics.gov.uk). The CPI captures the price of a basket of goods, and is often tweaked to take into account the change in consumer purchases over time. Sometimes products are dropped from the index and sometimes new products replace them, for example, the price of flights replacing the price of hotels in Margate. Other bodies such as the Organization for Economic Cooperation and Development (OECD) produce figures that are internationally comparable (that is, defined on the same basis across countries).

In the following sections, we show you how this process works by creating a market basket, seeing how this basket can be used to measure inflation and normalising it to a given base year so that calculating inflation rates between any two years is a piece of very agreeable cake.

Creating your very own market basket



The Consumer Price Index involves a large number of products and services – and is a big market basket. Understanding price indexes is easier if you create a simplified index with a very small market basket. In this section, we look at a *very* small market basket containing pizza, beer and textbooks. Because these three items are typical purchases of the undergraduate student population, we shall call it the Undergraduate Price Index.

For each of the three items in the Undergraduate Price Index, we create prices for 2007, 2008 and 2009 and list them in Table 5-1.

<i>Item</i>	<i>Number Bought</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>
Pizza	10	£10	£9	£9
Beer	60	£2	£2	£2.25
Textbooks	1	£120	£160	£170

In 2007, one medium cheese pizza costs £10, a pint of subsidised student union brown ale £2 and an overly long, poorly written, incomprehensible introductory economics textbook costs £120. The next year, the price of a medium cheese pizza actually falls to £9 because a new pizza outlet opens up

next to the old one, causing a price war. Beer still costs £2, but the university bookshop decides that it can really take advantage of the students, raising the price of the textbook to £160. (Don't worry about the 2009 column yet. We give you a chance to dig in and calculate inflation using the 2009 numbers later in the chapter.)

So far, so good. But in evaluating the index, you also have to keep track of how many of each item the typical student buys each year. For the sake of simplicity, assume that a typical student buys ten cheese pizzas, 60 beers and one economics textbook each year.

Calculating the inflation rate

To calculate how much inflation your university economy has (or deflation, if the cost of living happens to go down), first total up how much the market basket costs each year. In 2007, it costs £340: £100 on pizza (ten pizzas at £10 each), £120 on beer (60 beers at £2 each), and £120 on economics textbooks (one textbook at £120). The cost of buying the same market basket in 2008 is £370. So the cost of buying the same market basket has gone up by £30.

Now that you've done the adding, you need to do some simple algebra. Economists use the capital letter P to denote how many pounds the defined market basket costs. So in this case, P_{2007} means the cost of buying the market basket in 2007 and P_{2008} is the cost of buying the market basket in 2008. Because P as a letter is now in use, we're going to use pi, the Greek letter π (pronounced 'pie') as shorthand for the rate of inflation.

To calculate the rate of inflation, you use a very simple formula:

$$\pi = (P_{\text{Second Year}} - P_{\text{First Year}}) / P_{\text{First Year}} \quad (1)$$

In this case, the formula becomes:

$$\pi = (P_{2008} - P_{2007}) / P_{2007} \quad (2)$$

Substituting in $P_{2007} = £340$ and $P_{2008} = £370$, you find that $\pi = 0.088$. Multiply by 100 to convert this number into a percentage, and inflation in the Undergraduate Price Index is 8.8 per cent between 2007 and 2008. So, on the basis of this number, a student needs 8.8 per cent more money in 2008 to buy the simple market basket.

Setting up a price index

The undergraduate market basket is a simple example, but when government statisticians compute the Consumer Price Index, they basically do the same thing, just using a lot more goods. They also introduce the concept of a *price index* (or *price level index*) to make calculating and interpreting inflation rates over several years much easier. To set up a price index, they first establish a base year, or index year. Continuing our example, suppose that 2007 is the base year for the Undergraduate Price Index. You can then make a handy mathematical transformation so that the price level in 2007 is fixed at the number 100 and the price levels of every other year are set up so that they're relative to the 100 of the base year.

To make $P_{2007} = \text{£}340$ your base year, divide it by itself. That, of course, gives you 1, which you then multiply by 100 to get 100 ($100 \times 1 = 100$). This may seem like an idiotic thing to do until you realise that if you do the same thing to the other years, you end up with something very useful. Divide P_{2008} by P_{2007} and then multiply that product by 100 to get 108.8. This number is easy to interpret: it's 8.8 per cent larger than 100. Or, put differently, the price level in 2008 is 8.8 per cent larger than the price level in 2007. (Of course, you already discovered this inflation rate using equation (1) in the previous section.)

You can keep going, using the numbers for 2009 that appear in Table 5-1. For example, $P_{2009} = \text{£}395$. If you divide P_{2009} by P_{2007} and multiply by 100, you get 116.2; the price level in 2009 is 16.2 per cent bigger than the price level in 2007.

Working out the rate of inflation between 2008 and 2009 using these index numbers is also easy. Because the price index level for 2008 is 108.8 and the price index level for 2009 is 116.2, inflation is simply $(116.2 - 108.8) / 108.8 = 0.068$, or 6.8 per cent. (You're using equation (1) here, but you're inputting index numbers instead of actual costs of market baskets.)

Figure 5-1 charts the actual values of the Retail Price Index from 1987 to 2006. The index was set to a level of 100 using prices that consumers paid on average over the two-year period ending January 1987.

You can see that the Retail Price Index grew from its initial level of 100 in 1987 to a level of 200 in 2006. That is, to buy what a typical household consumes, you would have needed double the money in 2006 compared to what was needed in 1987, with the worst of the increases occurring in the late 1980s and early 1990s. Or to put it another way, increases in the money supply drove prices to double over this 20-year period.

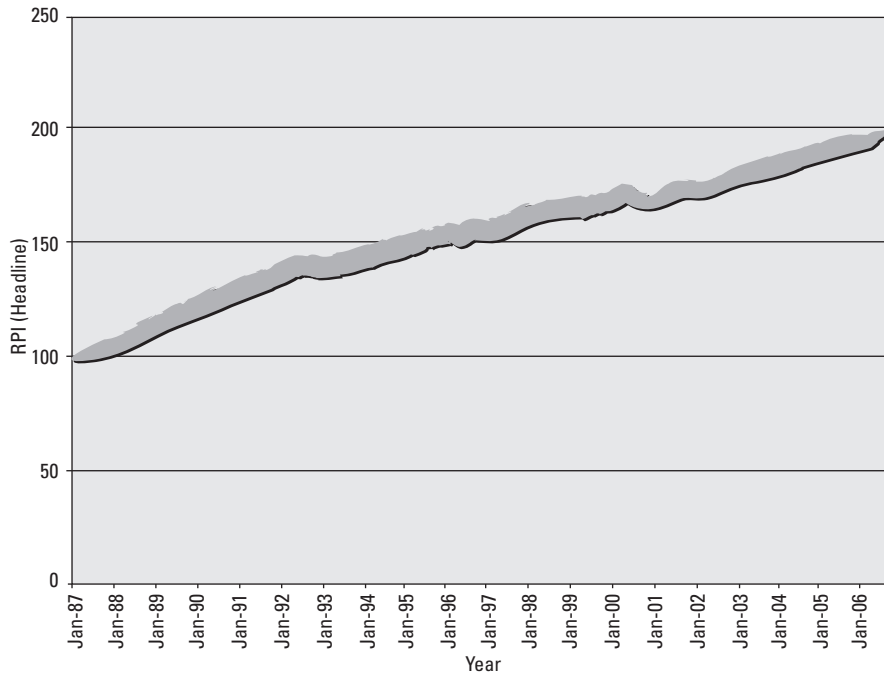


Figure 5-1:
Retail Price
Index,
1987–2006.

Determining the real standard of living with the price index



Beyond making inflation easy to measure and interpret, price indexes also make it simple to measure the very important difference between real prices and nominal prices. *Nominal prices* are simply money prices, which can change over time due to inflation. Because nominal prices can change, economists like to focus on *real prices*, which keep track of how much of one kind of stuff you have to give up to get another kind of stuff, no matter what happens to nominal prices.

For example, suppose that in 2008 you make £10 an hour working at a youth camp and the cost of a DVD is £20. The *real cost* of a DVD to you is two hours of work. Suppose that the next year, the prices of all goods double, but your wages also double so that you are earning £20 an hour and a DVD costs £40. The result is that you still have to work two hours to buy a DVD. So although the *nominal price* of a DVD has doubled, its *real price* in terms of labour – how much labour you have to give up to get a DVD – hasn't changed.

By constructing price indexes such as the CPI, economists can tell how the *real standard of living* changes for people from year to year. In the example of the previous section (using data from Table 5-1), inflation is 8.8 per cent between 2005 and 2006, meaning that the cost of living of a typical undergraduate student went up 8.8 per cent. So if at the same time student incomes go up only 5 per cent, students are actually worse off because costs have gone up faster than incomes. Real living standards – living standards measured in terms of how much stuff you can buy with your income – have fallen.

Identifying price index problems

Using price indexes to track the cost of living isn't a flawless system. Here are three big issues:

- ✔ **The market basket can never perfectly reflect family spending.** National Statistics tries to keep track of what a typical family of four purchases when calculating the Consumer Price Index (CPI). But families differ greatly, not only in terms of *what* they buy, but also in terms of *how many* of each thing they buy.
- ✔ **The market basket becomes outdated.** When to replace one product with another in the list is a fine judgement call. Suppose, for example, that National Statistics was deciding whether to include in the market basket DVD players instead of older VCRs. If Britain's national statisticians wait too long, they aren't capturing the change in purchases, but if they do it too soon, they may find that DVDs don't actually catch on and join the museum of dead technology (right next to the Sinclair ZX81 and the Betamax), in which case the figures would also be wrong.
- ✔ **The market basket can't account for quality.** Price isn't the only thing that matters to consumers. For example, what if a beer stays the same price but improves in quality from one year to the next? You're getting better beer for the same price, but this quality isn't reflected in the data. This problem is especially severe for things like computers, mobile phones and video games. For these products, quality improves dramatically year after year while prices stay the same or go down.

Each of these problems troubles government statisticians, who are constantly coming up with better price indexes and statistical methods to try to overcome them. The Federal Reserve Bank (the US government agency charged with determining the money supply) has recently come out with an estimate suggesting that the US CPI overstates inflation by 1 to 2 percentage points per year. Most of the overstatement comes from the failure of the CPI to account for new goods and quality improvements.

The main consequence of this overstatement is that the US government is overly generous with the cost-of-living increases it grants workers and retirees. Each year, government workers and retirees receive pay increases based upon increases in the CPI. These pay increases are designed to ensure that people's real incomes aren't eroded by inflation, but because the CPI is most likely overstating the rate of inflation each year, the cost-of-living increases are overly generous.

In the UK, regular revisions in the RPI and regular requirements for different presentations of the data have driven economists to distraction. However, in this case revisions are a necessary evil, because without them you can't easily compare prices historically.

Pricing the Future: Nominal and Real Interest Rates

Because inflation erodes the value of a loan repayment (see 'Tallying up the effects of inflation', earlier in this chapter), economists have to distinguish between *nominal interest rates* and *real interest rates*. Nominal interest rates are simply the normal money interest rates that you're used to dealing with; they measure the returns to a loan in terms of money borrowed and money returned. Real interest rates, however, compensate for inflation by measuring the returns to a loan in terms of units of stuff lent and units of stuff returned. This distinction is very important because what makes people want to save and invest is the *real* interest rate. After all, what lenders really care about isn't how much money they get back but how much stuff they can buy with it.

Suppose that you borrow £1,000 with the promise to pay £1,100 to the lender in a year. Your nominal interest rate is 10 per cent because you're paying back an additional £100, or 10 per cent more pounds than you borrowed. But if inflation occurs, the amount of stuff that £100 can buy decreases over time.



Say a nice meal for two with a bottle of wine costs £100 right now but is going to cost £105 next year. Right now, the lender is giving up 10 of these very good meals (£1,000 divided by £100 per meal) in order to give you the loan. Next year, when she gets repaid £1,100, she can buy 10.47 meals at the price of £105. The lender is giving up 10 meals now in exchange for 10.47 meals next year, meaning that the real rate of interest on the loan is 4.7 per cent. Because of inflation, the real rate of interest on the loan is substantially less than the nominal rate.

When lenders and borrowers negotiate a nominal interest rate on a loan, they both try to estimate what the inflation rate is going to be over the period of the loan. This expected rate of inflation is denoted algebraically as π^e . (Don't confuse expected inflation, π^e , with actual inflation, π . The former is what people expect to happen ahead of time, whereas the latter is what actually ends up happening.) The following sections show you how to estimate and use this rate.

Using the Fisher equation

Economist Irving Fisher came up with a simple formula, known as the *Fisher equation*, which links nominal and real interest rates. Using i to denote the nominal interest rate and r to denote the real interest rate, the formula is as follows:

$$i = r + \pi^e \tag{3}$$

This equation simply says that the nominal interest rate is the real interest rate plus the expected rate of inflation. This relationship is very important to borrowers and lenders because although all loan contracts specify a nominal rate of interest, their goal is to achieve a specific real rate of interest, even after any subsequent inflation reduces the value of money. By using the Fisher equation, the borrowers and lenders can determine what nominal interest to charge now in order to achieve a given real rate of return, taking into account the expected rate of inflation.

To see how this works, suppose that a borrower and lender agree that 6 per cent is a fair real rate of interest, and they also agree that inflation is likely to be 3.3 per cent over the course of one year. Using the Fisher equation, they write the loan contract with a 9.3 per cent nominal interest rate. A year later, when the borrower repays the lender 9.3 per cent more money than was borrowed, that money is expected to have only 6 per cent more purchasing power than the borrowed money, given the expected increase in prices.

Realising that predictions aren't perfect

Negotiations of the type described in the previous section depend crucially upon estimating the expected inflation rate, π^e , and lots of economists' job descriptions consist primarily of trying to predict future inflation rates. Their predictions are widely reported in the business media, but every person comes up with his or her own inflation forecast in an individual way. Some

people listen to the experts, whereas others make estimates based on their own daily experiences.

Note, though, that because forecasts aren't 100 per cent accurate, no one can say for sure what the real rate of return on the loan is going to be. For example, if the inflation rate turns out to be 9.3 per cent in the previous example, the real rate of return is 0 per cent. On the other hand, if the rate of inflation is 0 per cent, the lender gets back 9.3 per cent more money and can buy 9.3 per cent more stuff, meaning a real rate of return of 9.3 per cent. (See the earlier section, 'Feeling printing press pressures: The politics of inflation', for a discussion of why borrowers like inflation and lenders don't.)



Figure 5-2 plots actual inflation rates along with average expected inflation rates from a US study. The actual rates come from the monthly US CPI numbers, and the expected inflation rates come from a poll of consumers taken every month by the University of Michigan. Actual inflation between January 1980 and January 1981 was about 13 per cent. By comparison, consumers who were asked in January 1980 what they thought the rate of inflation would be over the next 12 months, on average, told researchers that they expected about a 10 per cent inflation rate. So in that particular instance, the inflationary expectations of typical consumers were off by about 3 per cent.

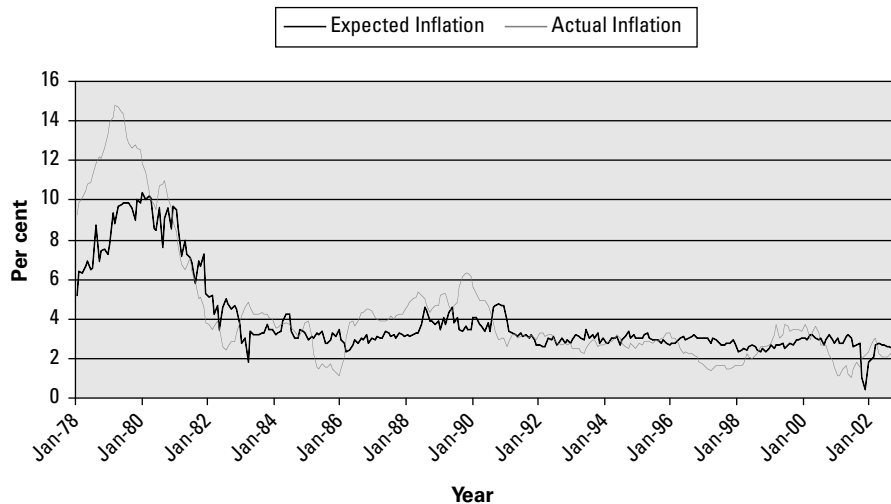


Figure 5-2:
Expected
and actual
inflation
rates.

Since about 1980, the two sets of numbers have been remarkably close, meaning that people's guesses about inflation in the past two decades have usually been wrong by no more than about 1 per cent. Of course, this period also corresponds to a period in US history where the government was committed to

low and stable inflation rates. Where governments are less committed to this aim, the expected and actual inflation are likely to diverge, because guesses are less likely to be correct. You can see in the US case shown in Figure 5-2 that the two sets of numbers differ the most in the period of highest inflation in the late 1970s. The predictability gained from following a low and stable inflation policy is generally believed to help people and businesses make good guesses on their investment decisions, and therefore enable them to plan more effectively. And because they can plan more effectively, people and businesses are more likely to increase their levels of investment. This reduced level of predictability, planning and investment is one of the reasons why Britain is held to have had a poorer investment performance than, for example, Germany (which has been more committed to low and stable inflation rates).

Chapter 6

Understanding Why Recessions Happen

In This Chapter

- ▶ Visualising the business cycle
 - ▶ Hoping for the ideal: letting price adjustments eliminate recessions
 - ▶ Dealing with reality: coping with sticky prices and lingering recessions
 - ▶ Linking slow price adjustments to slow wage adjustments
 - ▶ Introducing the Keynesian model
-

The biggest task of macroeconomists is to try to prevent – or at least shorten – *recessions*, those periods of time during which the economy's output of goods and services declines. Economists, politicians and most other people who work for a living despise recessions because of the high toll they exact in human suffering. When output falls, firms need fewer workers, and the typical result is massive layoffs, which cause significant increases in unemployment. Recessions can have long-lasting consequences: the recession of the 1980s was accompanied by lasting unemployment as UK manufacturing suffered, and it took several years for the housing market to recover in the 1990s. Anticipating and preventing or at least lessening those consequences is one of the tasks currently occupying policy makers following the most recent recession. In this chapter, we use the *aggregate supply/aggregate demand* model to show you how economists analyse recessions. Typically, recessions begin with what economists like to call *shocks* – unexpected adverse events such as terrorist attacks, natural disasters, the introduction of bad government policies or sudden spikes in the cost of important natural resources like oil.

The first big lesson of this chapter is that if the prices of goods and services in the economy were free to adjust to changes in demand and supply caused by shocks, the economy would typically be able to recover quite swiftly. Unfortunately, however, the second big lesson is that not all real-world prices

are totally free to adjust to shocks. Instead, some very important prices are quite slow to adjust – they are, as economists like to say, *sticky*. As a result, recessions can linger and cause a lot of harm unless the government intervenes to help the economy recover more quickly. (In Chapter 7, we discuss the best ways for governments to intervene.)

Throughout this chapter we use shocks to mean *negative* shocks – things and events that impact adversely on the economy. Bear in mind that shocks can be *positive* as well, that is, factors leading to sudden growth in demand in the economy such as the discovery of a new resource or a sudden and substantial tax cut.

Examining the Business Cycle

Economies go through alternating periods during which the output of goods and services expands and then contracts. In Chapter 4, we explain that Y represents the total output of an economy, so we use Y in this section to conserve some words.

The alternating pattern of economic expansion and contraction, which is illustrated in Figure 6-1, is often called the *business cycle* because businesses are so greatly affected by the changes in output.

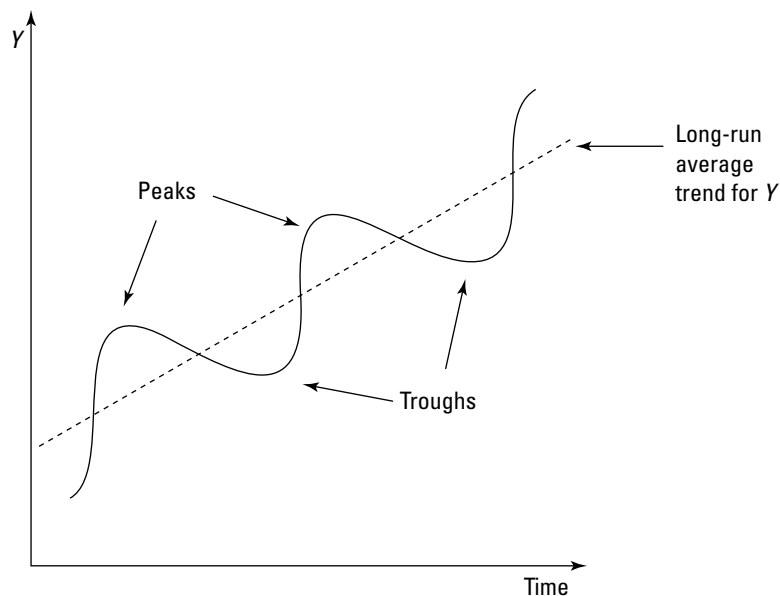


Figure 6-1:
The business cycle.



The solid line in Figure 6-1 represents how output, Y , varies over time. It alternates between troughs and peaks, which helps us identify periods of recession and recovery. Here's how we distinguish between the two:

- ✓ *Recessions*, or contractions, are the periods of time during which Y falls – that is, after a peak and before the next trough.
- ✓ *Recoveries*, or expansions, are the periods of time during which Y increases – that is, after a trough and before the next peak.

The dotted line in Figure 6-1 represents the long-run, average growth trend for Y . We draw Figure 6-1 with an upward sloping average growth trend for Y , to capture the fact that the economies of most countries now have sustained economic growth. In other words, on average, output tends to rise year after year. Because recessions still happen, however, the actual path of Y given by the solid line fluctuates around the long-run growth path given by the dotted line.

Looking at Figure 6-1, you can see that macroeconomic policy has two very natural goals:

- ✓ **Making the long-run average growth line as steep as possible:** The steeper it is, the faster (on average) output and living standards rise.
- ✓ **Reducing the size of business-cycle fluctuations around the long-run average growth line.** Smaller distances between peaks and troughs translate into fewer people suffering through bouts of unemployment when output falls.

In Chapter 7, we explain the policies that economists think are best for achieving these two goals. But in order for Chapter 7 to make sense, we must first explain what causes the business cycle – especially recessions and the high rates of unemployment that accompany them. After all, if you don't understand what's wrong, you can't sensibly fix it.

Striving for Full-Employment Output

Before you can say whether an economy is doing well or doing poorly, you need some objective standard of what 'doing well' is. Economists use the concept of *full-employment output* (which is represented by the symbol Y^*) as their measure of how well an economy should be doing.



The idea of full-employment output revolves around the concept of *full employment*, by which economists mean a situation in which everyone who wants a full-time job can get one. *Full-employment output* is how much output is produced in the economy when full employment exists in the labour market.

Please don't confuse full-employment output with the economy's *maximum output*, which is the larger amount of output that would be produced if everyone were forced to work as much as humanly possible.

Also, don't make the mistake of thinking that full employment is the same thing as having a zero unemployment rate. Even when everyone who wants a job can get one, some unemployment always exists as people voluntarily leave one job to search for a better job. For the duration of their job search, these people are counted as unemployed. Similarly, although the economy may be growing, *some* firms may be laying off workers, and those workers may be out looking for jobs. Economists call this situation *frictional unemployment*, as though the delay in finding a better job is due to some sort of friction slowing the process down.

As technology improves, full-employment output (Y^*) grows because better technology means that a fully employed labour force can produce more output. But to simplify their analyses, economists usually ignore the long-term growth trend and look only at whether actual output, Y , is currently above or below their best estimates of Y^* at that particular moment.

We follow this convention too for the rest of the chapter. Consequently, you're going to find out how the economy adjusts to situations in which output is above or below potential output at a given point in time.

As we show you in this chapter, the economy naturally wants to adjust back to Y^* anytime it deviates from Y^* . If that adjustment process was rapid enough, you wouldn't have to worry about business cycles, recessions and unemployment. If the economy reverted back to Y^* fast enough, recessions would be too brief to cause any serious negative consequences. Unfortunately, the natural adjustment process can be very slow, and as a result, recessions can be quite lengthy and awful.

Returning to Y^ : The Natural Result of Price Adjustments*

After an economic shock, such as a natural disaster or a spike in the cost of natural resources, price adjustments tend to return an economy to producing at full-employment output (Y^*). That's right, we said *price adjustments* – not the actions of government, the Chancellor of the Exchequer, the Governor of the Bank of England or even the European Central Bank. Don't believe us? Read on.

Consider a situation in which the *aggregate* (total) demand for goods and services in the economy falls off: individuals, firms and the government demand and buy less output than the economy is currently producing. The result is an excess supply of output which, in turn, leads to lower prices. After all, what does any business do when it can't sell off all its products at the prices it's currently charging? It lowers prices and has a sale. The lower prices attract more buyers, and soon the business can sell off the rest of its output.

This process repeats itself all over the economy during an economic downturn. When aggregate demand falls off due to an economic shock, firms lower prices to make sure that they sell off their outputs. This process eventually leads to two outcomes:

- ✓ Prices all over the economy fall (more or less).
- ✓ The economy again produces at full-employment output, Y^* .

For this process to work well, prices must be able to change quickly; if they can, the economy very quickly returns to Y^* . If, however, price adjustments are slow, the economy may produce less output than Y^* for a significant amount of time. In other words, if prices don't adjust quickly, you can get a recession. And until prices do adjust, the recession lingers.

We've just given you the briefest overview possible of how the economy responds to an economic shock. The next section provides much more detail so that you can understand how and why the economy eventually gets back to Y^* (and so you can pass your next exam, if that's your goal).

Responding to Economic Shocks: Short-Run and Long-Run Effects

Economists like to break the time period after an economic shock into two parts, which they call the *short run* and the *long run*. An easy way to get a handle on what these two parts mean for the macroeconomy is with the following:

- ✓ The *short run* refers to the period of time in which firms haven't yet made price changes in response to an economic shock.
- ✓ The *long run* refers to the period of time after which firms have made all necessary price changes in response to an economic shock.

These definitions are intentionally vague because the speed at which firms adjust prices varies from shock to shock. In this section, we show you that major differences exist between what happens in the short run and the long run.

We use a simple model that *holds constant* the money supply. Bear in mind that as the model gets more complicated, we can do away with some of these constraints. To get the idea, start from the simple model and work your way up!

Defining some critical terms

To see the difference between an economy responding to a shock in the short run versus the long run, begin by looking at Figure 6-2, which is a model of the macroeconomy. The horizontal axis measures the value of the output of goods and services sold in the economy (Y). This number is the same as a country's gross domestic product (GDP), which we discuss in Chapter 4. The vertical axis measures the overall price level in the economy, P .

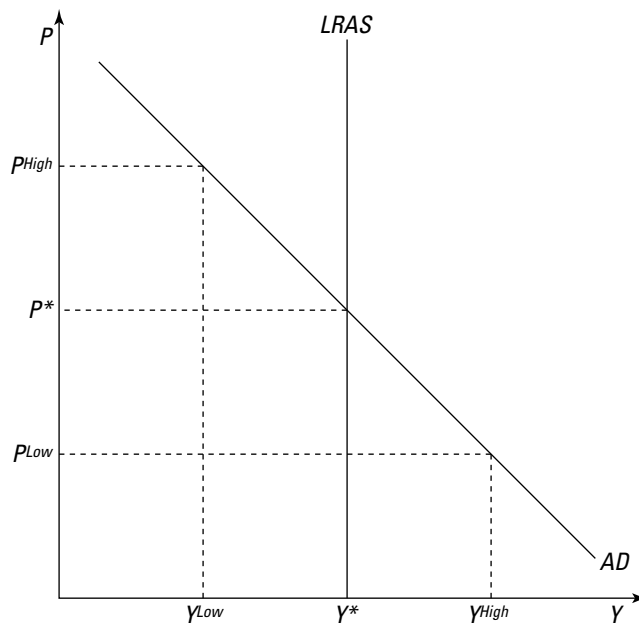


Figure 6-2:
A model of
a macro-
economy.

To understand the meaning of P , consider this: although each individual good and service has its own price, and some of those prices may be going up while others are going down, an overall trend in prices exists for the economy as a whole. P is simply a measure of how the prices of goods and services as a whole behave. If P goes up, on average prices are rising; if P goes down, on average prices are falling. And if prices stay the same, P (of course) stays the same. See Chapter 5 for details about how economists measure P .

In Figure 6-2, you see the symbol P^* . This symbol represents the *equilibrium* level of prices. What does that mean? P^* is the price level at which consumers want to buy exactly the amount of full-employment output (Y^*).

How do economists determine P^* ? That price level is determined by the intersection of what's called the *long-run aggregate supply curve* ($LRAS$) with the *aggregate demand curve* (AD). Before you start hyperventilating, we explain what these things are:



- ✓ The *aggregate demand curve* represents the total amount of goods and services that people want to buy.

Notice that in Figure 6-2, the AD curve slopes downward. That's because an inverse relationship exists between the price level and the amount of stuff that people want to buy. *Inverse relationship* simply means that at the higher price level (P^{High}), people want to buy a low level of output (Y^{Low}). But if prices fall to P^{Low} , people demand a much greater amount of output (Y^{High}). The downward slope of the AD curve captures the fact that at lower prices, people buy more.

- ✓ The *long-run aggregate supply curve* represents the amount of goods and services that an economy is going to produce when prices have adjusted after an economic shock.

In Figure 6-2, you can see that the $LRAS$ is a vertical line – it isn't a curve at all! (If you feel cheated, remember that a straight line is just a special type of curve, one without any curvature!) The $LRAS$ is drawn above the point on the horizontal axis that represents the full-employment output level, Y^* . Why? Because in the long run, changes in prices *always* return the economy to producing at the full-employment output level.

Still don't believe us? You are a sceptical bunch. We're going to convince you, but first you have to pay attention because here comes the science.

The Tao of P : Looking at price adjustments in the long run

Examine what happens if the economy starts out at a price level other than P^* . For example, look again at price level P^{High} and its corresponding aggregate demand level, Y^{Low} . Obviously, Y^{Low} is less than the economy's full-employment level of output (Y^*). That's important because firms would rather produce at output level Y^* . In fact, they've invested in factories and equipment that will be wasted if they produce at lower levels of output. Consequently, their response is to cut prices in order to increase sales. And they continue to cut prices until the overall price level in the economy falls down to P^* , because that's the price level at which consumers want to buy exactly Y^* worth of output.



Are you worried that all these price cuts will cause firms to lose money? Take heart: firms don't necessarily lose profits in this situation because their costs are falling at the same time. That's because when the economy is producing at less than Y^* , a lot of unemployed workers are available, as well as a lot of unused productive inputs such as iron and oil. Unemployment puts downward pressure on wages; in other words, having lots of labour readily available means you can hire people at lower wages. And the more piles of unused productive inputs that exist, the more their prices fall.

Okay, so the lower prices attract more customers, increase sales and cause the firms to hire back unemployed workers. This process continues until prices fall all the way to P^* , at which point the economy is operating at full employment again, meaning that all workers who want full-time jobs can get them.

In a similar fashion, prices can't remain below P^* for long. At price level P^{Low} , people want to buy Y^{High} worth of output. But that's more than firms can produce at full employment. The only way to produce that much output is if employees work longer than the standard working week. The only way to get them to do so is to pay them more, and the only way to give them higher wages is for firms to raise prices. So with demand exceeding supply, prices are raised until they reach P^* , at which price level the quantity demanded by consumers is exactly equal to the full-employment output level, Y^* .

As you can see, if prices have enough time to adjust, the economy always returns to producing at output level Y^* . Because we're calling the time required for prices to adjust the *long run*, it makes sense to call the vertical line above Y^* the *long-run aggregate supply curve*, because it shows how much output the economy will supply after prices have had enough time to adjust

to equalise the supply and demand for goods and services. (For much more about supply and demand, see Chapter 8.)

A shock to the system: Adjusting to a shift in aggregate demand

The previous section shows what happens if the prices of goods and services are, on the whole, too high or too low: they eventually adjust to the equilibrium price level (P^*), and so the economy can get back to producing at the full-employment output level (Y^*). But what causes the prices to be too high or too low in the first place? The usual cause is a shock to *aggregate demand* – the total amount of goods and services that people are willing to buy.

First, visualise what a shock to aggregate demand looks like: Figure 6-3 shows the aggregate demand curve shifting to the left from AD_0 to AD_1 . A leftward shift of aggregate demand is called a *negative demand shock*, and it may be caused, for example, by a decline in confidence in the economy that makes people want to save more and consume less. (A rightward shift of AD would be called a *positive demand shock*.)

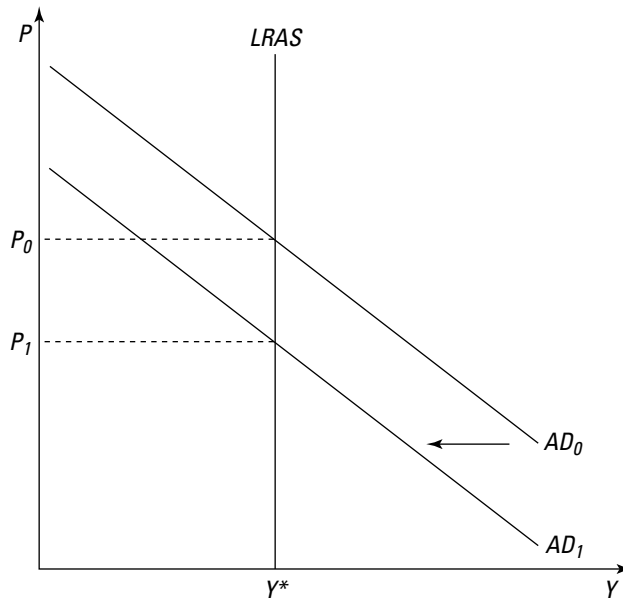


Figure 6-3:
A negative
shock to
aggregate
demand.

The original price level, P_o , was determined by where the original AD_o curve intersected the vertical $LRAS$ curve. In the long run, after firms adjust to the demand shock, the new price level, P_1 , is where the new AD_1 curve intersects the vertical $LRAS$ curve.

The new price level (P_1) is less than the original price level (P_o). Why? Demand for goods and services decreases after the negative demand shock. The only way to entice consumers to again purchase full-employment levels of output (Y^*) is to lower the cost of buying that much output, so the price level has to fall. Firms may take some time to make the necessary price reductions, but when they do, the economy again produces at Y^* in the long run.

We hope you're convinced by now that in the long run, after prices have a chance to adjust to whatever shocks occur, the economy again produces at the full-employment output level, Y^* . That's a huge contrast to what can happen in the short run before prices adjust, which we discuss next.

Dealing with fixed prices in the short run

As we discuss in the previous sections, after an economic shock happens, prices eventually adjust to return the economy to full-employment output (Y^*). However, this process may take some time because in the short run, prices are essentially fixed. Even the managers of the most nimble firms need some time to decide how much to cut prices. And some firms aren't quite as nimble.



Suppose that a firm has printed up catalogues listing the prices of the things it sells. This firm distributes catalogues only once a year, which means it is committed to selling to customers at these prices until the next catalogue is sent out. In such a situation, a firm adjusts its production to meet whatever amount of demand happens to come along at these fixed prices. If a lot of people show up to buy at these prices, the firm increases production, typically by hiring more employees. If very few people show up to buy, it reduces production, typically by hiring fewer employees.

Figure 6-4 depicts a situation in which firms have committed to a fixed set of prices and can respond to changes in demand only by adjusting their production levels. The figure shows the horizontal *short-run aggregate supply curve (SRAS)*, which is not a curve at all but a straight line. This 'curve' corresponds to price level P_o because the firms, in the short run, can't adjust their prices. Movements right and left along the *SRAS* curve capture the increases and decreases in output that firms have to make as demand for their products varies at the fixed price level.

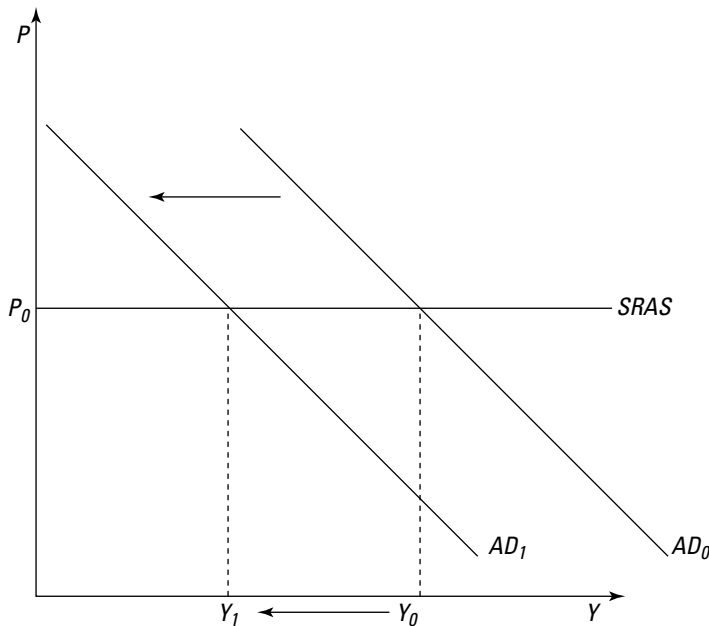


Figure 6-4:
The
short-run
aggregate
supply
curve.

In the short run, it makes sense to think of the firm as having more control over its production levels than its prices, which leads to modelling the *SRAS* curve as horizontal. In the long run, potential output is capped at a level given by a number of underlying long-run factors. In between, we have a number of models of various degrees of elaborateness explaining what happens. Generally at some point, these models all agree: the *SRAS* first becomes upward sloping (implying that prices and output rise together) and then vertical. What we're going to do here is to skip the middle bit in the name of simplicity. Of course, in reality economies don't just jump from a horizontal *SRAS* to a vertical *LRAS*, from growth to recession, or from one price level to another. However, the model is simpler to understand, without losing the essential features of the analysis, if we pretend for the moment that this does happen to economies. If you're okay with making that a deal, we're going to use only the initial horizontal curve and the final vertical curve, calling the former the *SRAS* and the latter the *LRAS*.

Figure 6-4 also has two aggregate demand curves, AD_0 and AD_1 , which again show what happens when aggregate demand is reduced as the result of a negative demand shock. The initial level of output that firms produce, Y_0 , is determined by the intersection of the original aggregate demand curve, AD_0 , with the *SRAS* curve. In other words, at price level P_0 , people demand output level Y_0 , and firms respond by supplying it.



Supermarkets and Y^*

The two most recent recessions in the United States, in 1991 and 2001, have been very mild – much milder than most previous recessions – and the early 1990s recession in particular was felt more strongly in other developed countries, such as the UK and Japan. The exact reason for this is not totally clear, but a number of economists have an interesting answer: inventory management. A combination of competitive pressure and technological progress enabled managers to manage prices more effectively than they did in the past. Inventory management means that stores typically carry very much less stock than they used to, order more flexibly and price more appropriately, meaning that stuff doesn't tend to hang about unsold for as long as it used to. Stocks in stores used to be counted by hand and priced by hand, usually once a month or so. Now, with the application of computerised systems, managers can adjust the prices of unsold stocks much faster. One of the poster boys for this revolution

is Wal-Mart, the US low-cost retailer and parent of UK supermarket chain ASDA, which has been the subject of countless business-school case studies on the subject. Wal-Mart has developed inventory management systems that are often acclaimed as the most sophisticated in the industry. With these computerised systems, Wal-Mart managers can tell minute by minute what's selling and what's not. As a result, the prices of slow-moving items are cut very quickly so that products don't go unsold for weeks or months. Of course, Wal-Mart's competitors quickly followed suit, and now these systems are commonplace within retail management.

As a result of such innovations, prices can adjust quickly to equate supply and demand. Prices can now fall much more rapidly to get the economy back to producing at full-employment output (Y^*). That means shorter, milder recessions.

When the negative demand shock strikes, it shifts aggregate demand leftward to AD_1 . Reduced demand means that at the fixed price level, customers are willing to buy less output. Because firms can't change prices, their only recourse is to reduce production down to match the decrease in demand; this reduced level of output (Y_1) appears on the graph where the $SRAS$ curve intersects AD_1 . Because lower output means that firms need fewer workers, you end up with a recession: output falls and unemployment rises.

If you compare Figures 6-3 and 6-4, you can see that the leftward shift in aggregate demand has very different effects in the short run and the long run:

- ✓ In the short run when prices are fixed, output falls and unemployment rises.
- ✓ In the long run, prices fall and output returns to the full-employment level.

Why the huge difference between the short run and the long run? Firms aren't forever stuck with their original catalogue prices. Eventually, they print new catalogues with lower prices. The lower prices entice customers to purchase more, and soon the economy can return to producing at the full-employment output level, Y^* .

Putting together the long and short of it

If you've got the previous sections tucked under your belt, you're now an expert in both long-run and short-run responses to an economic shock. (Congratulations! You've earned the right to pontificate at your next dinner party!) We now drive this subject home by putting the two very different responses together into one big picture.

Figure 6-5 lets you see how an economy adapts to a negative demand shock both in the short run and in the long run. The economy begins at point A , where the original aggregate demand curve, AD_0 , intersects both the $LRAS$ and the $SRAS$ curves. At point A , the economy is in equilibrium because at price level P_0 , the aggregate demand for output equals the full-employment level of output, Y^* . Neither a surplus nor a shortage exists that may cause prices to change.



The $SRAS$ curve is horizontal at price P_0 to reflect the fact that after the economy reaches its equilibrium (where AD_0 intersects the $LRAS$ at output level Y^*), the prices that are determined at that level are fixed in the short run; they can't change immediately even if a demand shock happens to come along.

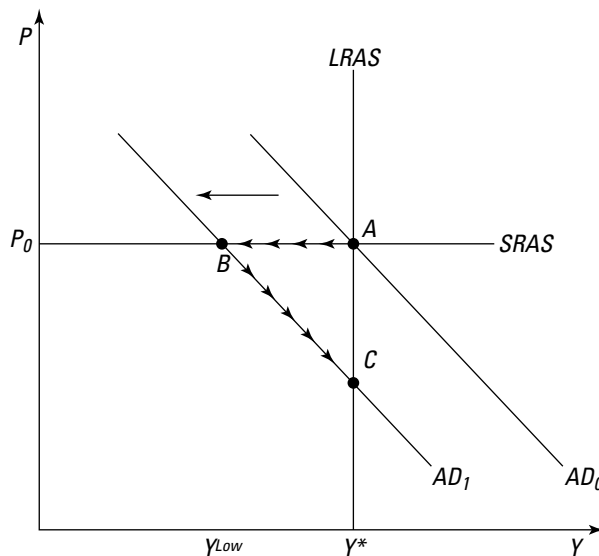


Figure 6-5: Short-run and long-run responses to a negative demand shock.

For example, suppose that the aggregate demand curve shifts left from AD_0 to AD_1 because of a negative demand shock of some sort. Because prices are fixed in the short run at P_0 , the economy's first response is to move from point A to point B . In other words, because prices are fixed, production falls from Y^* down to Y^{Low} as firms respond to decreased demand by cutting production. (Small arrows indicate the movement of the economy from point A to point B .)

At point B , the economy is operating below full employment, implying that there are a lot of unemployed workers. This high level of unemployment causes wages to fall. As wages fall, firms' costs also fall, allowing them to cut prices in order to attract more customers.

Falling prices cause increased aggregate demand for goods and services, which eventually moves the economy all the way from point B to point C . (This movement is indicated by arrows on the graph.) When the economy reaches point C , it is once again producing at full employment, Y^* .

The short-run and long-run effects of a negative demand shock are basically total opposites of each other:



- ✓ In the short run, prices are fixed while output decreases.
- ✓ In the long run, prices decrease while output returns to Y^* .

If prices don't stay fixed for very long, the economy can quickly move from A to B to C . But if prices are slow to adjust to the negative aggregate demand shock, the economy can take a very long time to get from A to B to C . In such cases, a long-lasting recession results during which output remains below Y^* and many people are unemployed.

For these reasons, we need to figure out what affects the ability of prices to change quickly. The most important culprit is sticky prices, or more precisely, sticky wages.

Heading toward Recession: Getting Stuck with Sticky Prices

When the economy encounters a negative demand shock like the one depicted in Figure 6-5, price flexibility (or lack of flexibility) determines both the severity and length of any recession that may result. If prices are infinitely flexible – if they can change within seconds or minutes after a shock – the economy immediately moves from point A to point C , and all is right with the world. But if prices are fixed for any period of time, the economy goes

into a recession as it moves from point *A* to point *B*, before prices eventually fall and bring it back to full-employment output at point *C*.

In the real world, prices are indeed somewhat slow to change, or, as economists like to say, prices are *sticky*. Interestingly, they tend to be stickier when going downward than upward, meaning that prices appear to have a harder time falling than rising.

The major culprit seems to be one particular price: wages. Wages are the price employers must pay workers for their labour. Unlike other prices in the economy, people are particularly emotionally attached to wages and how they change over time.

In particular, employees don't like to see their wages cut. They have a very strong sense of fairness when it comes to their wages and, as a result, usually retaliate against any wage cut by working less hard. (Not to mention that EU employment law has some pretty stern rules against firms cutting wages when workers have contracts!) As a result, managers typically find it counterproductive to lower wages even if a firm is losing money and needs to cut costs.

Cutting wages or cutting workers

Suppose that a negative demand shock hits an economy and greatly reduces sales at a particular company. The firm is losing money, so managers need to figure out a way to cut costs. About 70 per cent of this company's total costs are labour costs (wages and salaries). Naturally, labour costs are an obvious target for cuts.

But the managers of the firm realise that if they cut wages, employees will get angry and work less hard. In fact, their productivity may fall off so much that cutting wages may make the firm's profit situation worse: output may fall so much that sales revenues decrease by more than the reduction in labour costs. Therefore, cutting wages isn't really a good option.

So, instead, the managers lay off a large chunk of their workforce in order to reduce labour costs. For example, if sales are down 40 per cent, the firm may lay off 40 per cent of the workforce. However, any workers who remain employed get to keep their old wages so that they aren't angry and their productivity doesn't fall.



For the reasons we're showing you here, what you see during a recession is a large increase in unemployment but little decrease in wage rates. The fact that managers are unwilling to cut wages, however, has a nasty side effect: as we discuss in the next section, not cutting wages makes it very hard for firms to cut the prices of the goods and services they sell.

Adding up the costs of wages and profits

Obviously, firms need to turn a profit in order to stay in business. And that means making sure that the price per unit that they charge for their products exceeds the cost per unit of making them.

During a recession, lower aggregate demand means that firms reduce production and sell fewer units. As we discuss in the previous sections, wages are the largest component of most firms' costs – in fact, they're a full 70 per cent of the average firm's costs. If a firm can't cut wages for fear of causing worker productivity to drop, it also can't reduce its per-unit production costs very much. In turn, the firm can't cut its prices very much because prices have to stay above production costs if firms are to make a profit and stay in business.

What does all this mean? When demand drops off, prices are typically sticky. They stay high despite the fact that less demand exists for output in the economy. That's an underlying reason for the economy moving horizontally from point *A* to point *B* in Figure 6-5 after the negative demand shock. With prices sticky because firms can't cut wages, the negative demand shock results in a recession with output falling and unemployment rising because so many workers get fired.

Worse yet, unless prices can somehow begin to fall, the economy isn't able to move from *B* to *C* to get back to producing at the full-employment output level (Y^*). Prices do *eventually* fall, but this process can take a long time, meaning that the negative demand shock can cause a long-lasting recession.

Returning to Y^ with and without government intervention*

In Chapter 7, we explain how the government can use monetary and fiscal stimuli to get around the sticky prices problem by boosting aggregate demand. Here, we want to give you a preview of how that process works.



Imagine that after the negative demand shock depicted in Figure 6-5 moves aggregate demand leftward from AD_0 to AD_1 , the government doesn't hang around waiting for prices to eventually fall. Instead, it stimulates aggregated demand so that the aggregate demand curve shifts back rightward and returns to where it started, at AD_0 . Taking this action returns the economy to producing at full employment without having to wait for prices to fall.

What if the government doesn't act to stimulate aggregate demand in that fashion? What if the economy is at point *B* and the government doesn't intervene? In such cases, prices do eventually fall because firms' production costs eventually fall.

As we see in the previous sections, labour costs are very slow to fall because managers don't want to risk alienating workers by cutting their wages. But because there are so many unemployed workers when the economy is at point *B*, wages eventually decline. Some firms hire unemployed people at lower wages, which reduces their costs, meaning that they can undersell firms that keep wages high. Eventually, such competitive pressures mean that all firms end up cutting wages.

Other costs also decline, because during a recession, with output so much diminished, a significant portion of the economy's productive capacity is unused: unused factories, unused trucks, unused train cars and unused ships, as well as large amounts of unused lumber, iron, oil and other productive inputs.

The owners of these unused inputs lower their prices in order to try to sell them. As their prices fall, firm costs also fall, thereby allowing firms to reduce the selling prices of their output. And as these selling prices fall, the economy moves from point *B* to point *C* in Figure 6-5, restoring the economy to producing at the full-employment output level (Y^*). See how nicely it all (eventually) works out?

Achieving Equilibrium with Sticky Prices: The Keynesian Model

Even if this book is the first on economics you've ever laid your hands on, you may have heard the name Keynes before. Who is this guy, and why do economists like him so much?



John Maynard Keynes was the most influential economist of the 20th century. He was the first economist to realise that sticky prices (caused by sticky wages) are the culprit behind recessions. If you read the previous section, you may not have thought the ideas contained there were revolutionary, but trust us: Keynes's insight changed the way people studied economies.

What inspired Keynes to have this insight? He was led to the idea by the horrible state that the economy reached during the Great Depression of the 1930s. Just the name itself – *Great Depression* – gives you some idea how bad things got. Normal economic downturns are called *recessions*. Really bad

recessions are called *depressions*. (Yes, technical definitions apply to the two words – see the ‘What makes a recession a recession?’ sidebar, later in this chapter.) But what happened in the 1930s was so bad that people started calling it the *Great Depression* to indicate just how severe it was.

The Great Depression started with a lingering recession from 1929 to 1933. This recession is most usually described as beginning with the infamous stock market crash of 1929 (the Wall Street Crash). The United States, for example, didn’t see its output return to its 1929 level until after entering the Second World War in 1941, and the results for the UK were hardly better, although the recession of 1921 was actually the more severe. To put the Great Depression in perspective, look at Table 6-1, which gives data for each of the seven recessions that the United States has experienced since 1960, plus (on the first line) the same data for the Great Depression. We use US data because its consistency over time makes seeing a clear picture a little easier: you can also plot a similar picture for the UK or many other countries.

Table 6-1

**The Great Depression and US
Recessions since 1960**

<i>Start</i>	<i>End</i>	<i>Duration (Months)</i>	<i>Highest Unemployment Rate</i>	<i>Change in Real GDP (%)</i>
8/1929	3/1933	43	24.9	-28.8
4/1960	2/1961	10	6.7	2.3
12/1969	11/1970	11	5.9	0.1
11/1973	3/1975	16	8.5	1.1
1/1980	7/1980	6	7.6	-0.3
6/1981	11/1982	16	9.7	-2.1
6/1990	3/1991	8	7.5	-0.9
3/2001	11/2001	8	6.0	0.5

Source: NBER, Economic Report of the President, Bureau of Labor Statistics

As you can see, the Great Depression was far, far worse than any normal recession. Nearly 25 per cent of the labour force was unemployed, and the initial downturn lasted about four times longer than the 10.7-month average duration of post-1960 recessions.

Total economic output as measured by real GDP (which we discuss in Chapter 5) also fell much more than in a normal recession. Because real GDP adjusts for inflation, it captures changes in the physical quantity of output produced. In recent recessions, output has fallen at most 2 or 3 percentage points. During the Great Depression, it fell 28.8 per cent!

What makes a recession a recession?

At the beginning of the chapter, we define a *recession* as a period of time during which output falls and unemployment rises. But this definition isn't the only one. For example, you may read in a textbook or a newspaper article that an economy is in a recession if real GDP falls for two consecutive quarters. But if you look at Table 6-1, you notice that during certain recessions (like the one in the US that began in April 1960), real output actually went up rather than down. So why was that time period labelled a recession?

A lot of factors go into determining what gets called a recession. The best internationally accepted definition comes from the National Bureau of Economic Research (NBER) in the US, which is accepted and used by other bodies, including the Organisation for Economic

Co-operation and Development (OECD), although, for some obscure reason, governments tend to like using their own definition. NBER has a long set of criteria that begins with output falling and unemployment rising and includes lots of other things, such as how fast factories receive new orders. Sometimes these other factors cause the NBER to feel that the economy has passed a peak and has entered a recession even if output isn't falling. A good rule of thumb, one usually accepted by commentators in the UK, however, is that a recession is characterised by falling output measured by GDP over two successive quarters. Anything worse is a depression. Economists' humour characterises the difference as follows: in a recession, your neighbour loses his or her job, whereas in a depression, you lose yours.

In comparison, the misery of the Great Depression reached a peak in the UK in 1932. The rate of unemployment topped out at 22 per cent, although output measured by GDP fell by only about 5 per cent over the preceding years. The next worst recessions, in the early 1980s and 1990s, by contrast, were characterised by persistent high unemployment, peaking at 10 per cent, but falls in GDP of only 3 per cent and 1.5 per cent each.

As a witness to the Great Depression, Keynes obviously wanted to figure out what caused such a drastic economic downturn – and what may prevent such devastation from happening again.

Adjusting inventories instead of prices

Not only did Keynes figure out that sticky prices cause recessions, but he also developed a hugely influential model that's still presented in many macroeconomics textbooks. This model is a small part of a larger approach to managing the macroeconomy that came to be called *Keynesianism* – an

approach that favoured large government interventions into the economy rather than the sort of *laissez-faire* policies of non-intervention preferred by other people. (For a discussion of the costs and benefits of having the government intervene in the economy, see Chapter 3.)

To be fair, we have to point out that Keynesianism has attracted a lot of critics and is not the be-all-end-all of macroeconomics. But the part of it we present here is not controversial. This aspect of Keynesianism explains how an economy adjusts to *equilibrium* – a place where aggregate supply matches aggregate demand – in the extreme short run after an economic shock when prices can't change at all.

Look back at Figure 6-4 for a moment. The Keynesian model elaborates on exactly how an economy moves from producing at output level Y_0 to producing at output level Y_1 when a shock to aggregate demand happens and prices are fixed at level P_0 .

Keynes's model focuses our attention on firms' inventories of goods that have been made but not yet sold. According to Keynes, changes in inventories guide firms to increase or decrease output during situations in which prices are sticky and can't serve as signals of what to do.

To see the novelty of Keynes's inventory idea, understand that if prices were able to change, prices (not inventories) would guide firm decisions about how much to produce:

- ✓ If prices are rising, a firm knows that its product is popular and that it should increase output.
- ✓ If prices are falling, the firm knows that the product isn't doing well and that it should probably cut output (and maybe get into another line of business!).

In an economy with fixed prices, however, firms need some other way of deciding whether to increase or decrease production. Keynes realised that the guiding force would be changes in inventories.

Keeping an eye on target inventory levels

Inventories are constantly turning over, with goods flowing both in and out. New production increases inventories, while new sales decrease inventories. The two factors interact to determine whether inventories are rising, falling or staying the same. For example, if new production equals new sales, inventory levels stay constant. If new production exceeds new sales, inventories rise.

The interaction of new production and new sales is important because each firm has a *target level* of inventories that it likes to keep on hand to meet

situations in which sales temporarily run faster than the firm can produce output. The target level is determined by the costs and benefits of having a bigger or smaller inventory on hand.

Having less inventory than the target level is dangerous because the firm may not be able to keep up with sales spikes. Having more inventory than the target level is wasteful because there's no point in having stuff sitting around unsold, year after year. Each firm weighs these costs and benefits to come up with its own target inventory level.

Target inventory levels may vary from year to year depending on whether firms are expecting strong or weak sales. If managers are expecting strong sales, they may plan on increasing inventories, whereas if they are expecting weak sales, they may plan on decreasing inventories.

Keynes realised that aggregate demand shocks (which are, by definition, unexpected) would show up as unexpected changes in firm inventories.

- ✔ Unexpectedly low aggregate demand means that sales slow so much that inventories increase and reach levels higher than firms had planned on.
- ✔ Unexpectedly high aggregate demand means that sales increase so much that inventories decrease and reach levels lower than firms had planned on.

Increasing or decreasing output as inventories fluctuate

Unexpectedly large changes in inventories cause firms to change their output levels as follows.

- ✔ If inventories rise above target levels, firms respond by cutting production. By reducing production rates to less than sales rates, inventories begin to fall down toward target levels.
- ✔ If inventories fall below target levels, firms respond by raising production. By increasing production rates to more than sales rates, inventories begin to rise toward target levels.

The changes in output levels caused by changes in inventories are hugely important because they determine not only whether output (Y) is increasing or decreasing, but also whether unemployment is rising or falling.

For example, if firms increase production because inventories have fallen below target levels, they need to hire more workers, and unemployment falls. If, on the other hand, firms decrease production because inventories rise above target levels, they need to lay off workers, and unemployment rises.

Adjusting inventories based on planned and actual expenditures

The Keynesian model differentiates between planned expenditures and actual expenditures as follows:

- ✓ *Planned expenditures* are the amount of money that households, businesses, the government and foreigners want to spend on domestically produced goods and services.
- ✓ *Actual expenditures* are equal to gross domestic product (GDP), which we discuss in Chapter 4; they are what households, businesses, the government and foreigners actually end up spending on domestically produced goods and services.



What happens when actual expenditures are different from planned expenditures? Inventories automatically change. For example, if more money is spent on goods and services than was planned, people are buying up more output than is currently being produced. This situation is possible because firms sell goods from their inventories that were produced in previous periods. On the flip side, if people spend less money on goods and services than was planned, firm inventories rise because firms have to store up all the output that they can't sell.

Keynes represented planned expenditures, PE , algebraically with the following equation:

$$PE = C + I^p + G + NX \quad (1)$$

What do all these letters mean? We discuss them in detail in Chapter 4, but here's the short version:

- ✓ C stands for the amount of output that consumers want to consume.
- ✓ I^p stands for the amount of output that firms plan to buy as investment goods, such as new factories and equipment, as well as any inventory changes that firms plan to make.
If, later on, firms have to increase or decrease inventories more than they planned, actual investment, I , doesn't equal planned investment, I^p .
- ✓ G stands for how much output the government wants to buy for things such as building schools or ensuring an adequate supply of paper for paperwork.
- ✓ NX stands for *net exports* – the value of our exports minus the value of our imports. NX tells us the net demand that the foreign sector of the economy has for stuff that we make domestically.

For actual expenditures, Y , Keynes used the same equation that we use to calculate gross domestic product (which we discuss in Chapter 4):

$$Y = C + I + G + NX \quad (2)$$

Why can we use the GDP equation to calculate actual expenditures? As we explain in Chapter 4, actual expenditure is equal to national income because every penny of expenditure made in the economy is income to somebody. Furthermore, actual expenditure is also equal to the pound value of all goods and services produced in the economy because every bit of output produced is sold to someone. (This process is actually part of the way that stock is valued for accounting purposes: any output that a firm makes but can't sell to customers is counted as being 'sold' by the firm to itself as that output is placed into inventory. These inventory changes are known as *inventory investment* and are totalled up in GDP as part of the total investment, I .)

Having three ways of looking at Y is actually very handy as you become familiar with the Keynesian model. Sometimes, understanding the model is easier if you think of Y as being actual expenditures; at other times understanding is easier if you think of Y as being national income or output. We switch between these three definitions whenever doing so helps make understanding the model easier.

The only difference between the right-hand sides of equation (1) and equation (2) is the investment variable, which is *planned investment* (I^p) in the first equation and *actual investment* (I) in the second. In other words, Y and PE differ only because of differences in investments caused by inventories increasing or decreasing unexpectedly when sales are more or less than planned.

Bringing some algebra into the mix

You knew the time was coming: here's where things get algebraic. Our goal? To identify the Keynesian model's economic equilibrium by using our mathematical superpowers. (Quick – to the nearest surviving public telephone box!)

First, we need to define a *consumption function* – a way to calculate total consumption – that we can substitute into equation (1). In Chapter 4, we present the following formula for calculating consumption:

$$C = C_o + c(1 - t)Y \quad (3)$$

For all the details, look back at Chapter 4. For now, what you really need to know about this formula is that higher income (Y) leads to higher consumption (C).

If you substitute equation (3) into equation (1), you get:

$$PE = C_o + c(1 - t)Y + P + G + NX \quad (4)$$

If you look carefully, you see that this equation shows that the total planned expenditure on goods and services in the economy (PE) depends on the total income in the economy (Y). The higher the total income, the more money people are going to plan to spend.

A good way to simplify this equation is to create a variable called A and to define it as follows:

$$A = C_o + P + G + NX$$

When you do that, equation (3) looks a little more palatable:

$$PE = A + c(1 - t)Y \quad (5)$$

The variable A stands for *autonomous expenditures*, by which economists mean the part of planned expenditures that doesn't depend on income (Y). The part of planned expenditures that does depend on income, $c(1 - t)Y$, is known as *induced expenditures*.

To understand induced expenditures, you need to realise that because t stands for the income tax rate, $(1 - t)Y$ is what people have left over to spend after the government taxes them. And of that amount, the fraction c gets spent on consumption, so that $c(1 - t)Y$ tells you how much expenditure is 'induced' by an income of size Y .

Figure 6-6 graphs equation (5) and labels it the *planned expenditure line*.

To find the specific equilibrium of the Keynesian model, understand that all possible equilibriums are captured by the following equation:

$$PE = Y \quad (6)$$

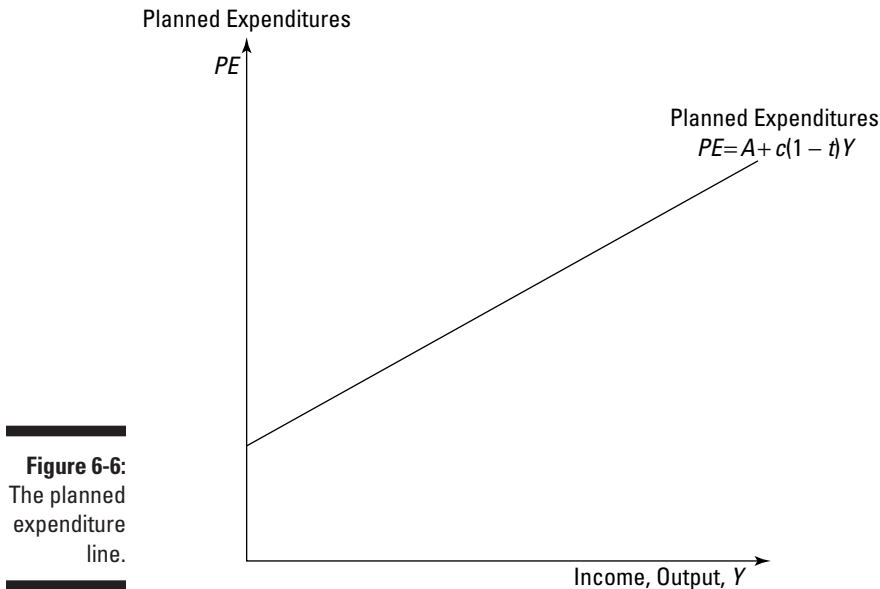


Figure 6-6:
The planned
expenditure
line.

This equation can be read as ‘planned expenditures equal actual expenditures’. (Remember that Y equals both total income and total expenditure in the economy because all expenditures are income to somebody.)

Any situation where $PE = Y$ is an equilibrium. Why? Because if the economy can get to the point where $PE = Y$, nobody has any reason to change behaviour. Consumers are consuming as much as they planned to consume (C). The government is buying up as much output as it wanted to buy (G). Foreigners are buying as much stuff from us as they intended (NX). And, most importantly, firms are spending exactly as much on investment as they planned – implying that inventories aren’t changing unexpectedly.

If planned expenditures equal actual expenditures, you truly have an equilibrium because everybody is getting what they want, and nobody has any incentive to change behaviour.

You can solve the equilibrium value of output, which we call \tilde{Y} , by substituting equation (5) into equation (6). If you do so, you get the following:

$$\tilde{Y} = A + c(1 - t)Y \quad (7)$$

Showing equilibrium graphically

If the last equation is just too frightening, stick with us because finding the Keynesian model's equilibrium graphically is much easier. To do so, you plot the $PE = Y$ equation on the same graph as the $PE = A + c(1 - t)Y$ equation, as we do in Figure 6-7. The point where the two lines cross is the equilibrium. At that point, planned expenditures exactly equal actual expenditures in the economy.

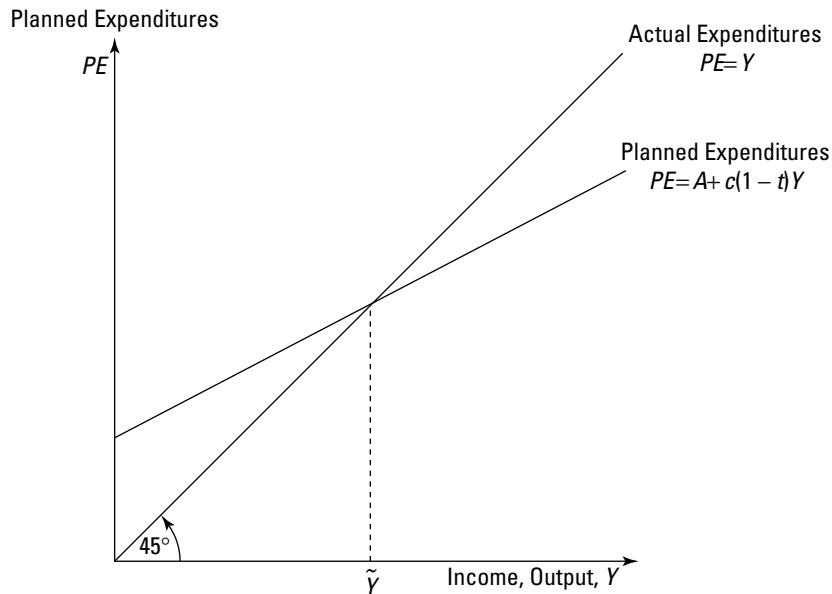


Figure 6-7:
The
Keynesian
model's
equi-
librium, \tilde{Y} .

This equilibrium is *stable*, by which we mean that if the economy starts out at any income level other than \tilde{Y} , it soon moves back to \tilde{Y} . The thing that returns the economy to \tilde{Y} is inventory changes.

To see why this is true, look at Figure 6-8, which exploits a nifty geometric trick about the $PE = Y$ line to show how the economy behaves when it's not producing at the equilibrium output level, \tilde{Y} .

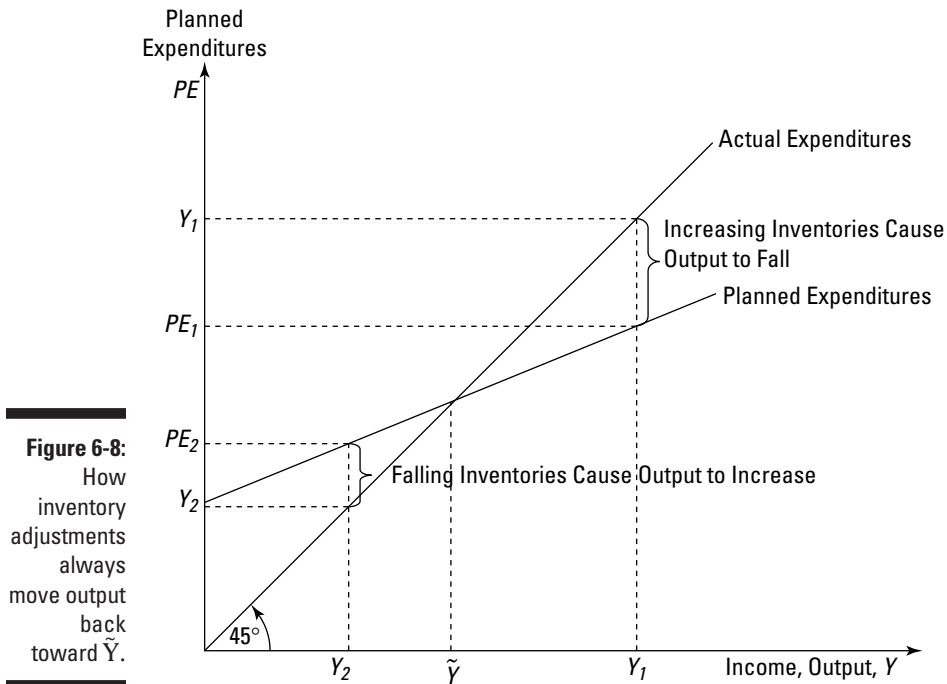


Figure 6-8:
How
inventory
adjustments
always
move output
back
toward \tilde{Y} .

The trick is that the $PE = Y$ line shows up on the graph at a 45-degree angle, meaning that it can be used to draw squares – shapes whose sides have the same length. That means you can transpose any value of Y onto the vertical axis. To do so, take any value of Y , go straight up until you hit the 45-degree line, and then go straight sideways until you hit the vertical axis. The point you hit represents as many pounds vertically as Y represents horizontally.

For example, in Figure 6-8, start on the horizontal axis at output level Y_2 , which is less than the equilibrium output level \tilde{Y} . If you go up vertically to the 45-degree line and then to the left, you can plot output level Y_2 onto the vertical axis. Why is this useful? Because Y_2 can then be compared directly with the level of planned expenditures, PE_2 , which you get by starting at output level Y_2 on the horizontal axis.

As you can see, $PE_2 > Y_2$, meaning that planned expenditures exceed output in the economy. This situation means that inventories are going to drop unexpectedly as firms sell part of their stockpiles of inventory to make up for the fact that people are buying up more stuff than firms are currently producing. This drop in inventories returns the economy to equilibrium.

As inventories fall unexpectedly, firms increase production. As a result, Y increases. Furthermore, Y continues to increase until it reaches \tilde{Y} because for any value of $Y < \tilde{Y}$, you can see from the graph that planned expenditures continue to exceed output.

Inventory adjustments also return the economy to equilibrium if it starts out at an output level like Y_j , which is greater than \tilde{Y} . As you can see in Figure 6-8, by using the 45-degree line, actual output, Y_j , exceeds planned expenditures, PE_j . In other words, people are buying less (PE_j) than firms are currently producing (Y_j), so inventories are going to start to rise.

Firms respond to increases in inventories by reducing output. They lay off workers and cut production. As a result, Y falls. Y continues to fall until it reaches \tilde{Y} because for any value of $Y > \tilde{Y}$, you can see from the graph that output is going to continue to exceed actual expenditures.

Boosting GDP in the Keynesian model

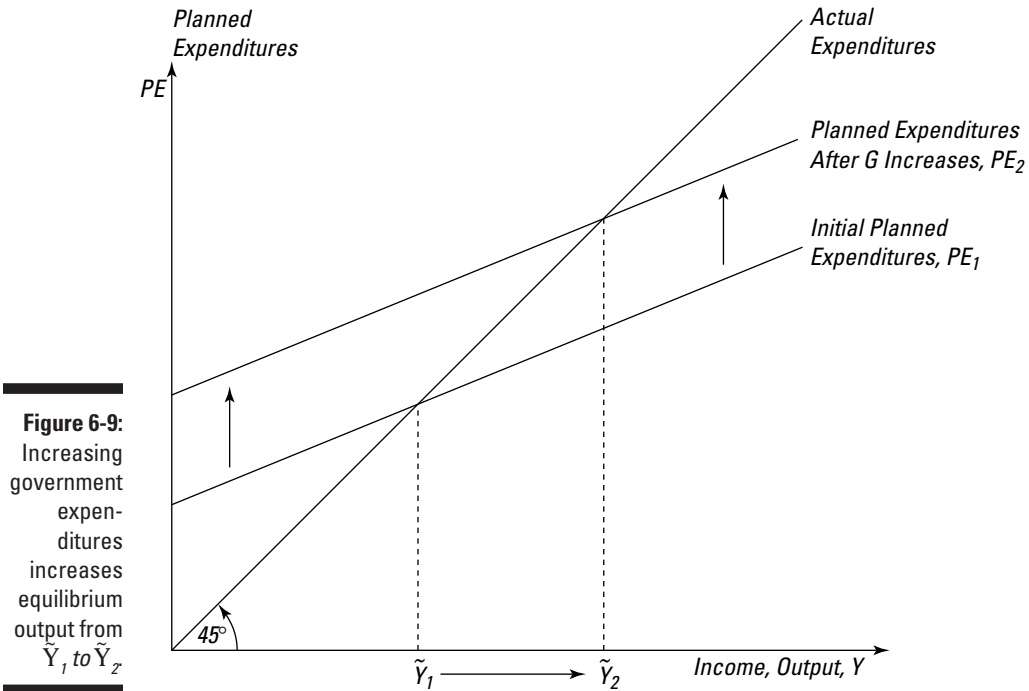
Keynes didn't just invent his model to explain how economies with sticky prices reach a stable equilibrium. What he really wanted to do was to use it to show what governments can do during a recession to make things better.

For example, consider Figure 6-8 once again. Suppose that inventory adjustments have carried the economy to equilibrium income, \tilde{Y} , but that \tilde{Y} is less than the economy's full-employment output level, Y^* . In such a case, Keynes asked, what – if anything – should governments do?



Governments can choose to do nothing. Eventually, because $\tilde{Y} < Y^*$, prices will fall and the economy will return to full employment (as it does moving from point B to point C in Figure 6-5). But Keynes argued that governments would be able to speed up the recovery by boosting planned expenditures.

For example, suppose that the government decides to increase G , government spending on goods and services. If it does so, PE in equation (4) clearly gets bigger. Because G is a part of autonomous expenditures (A), the increase in G means an increase in A in equation (5). Graphically, a larger A means that the planned expenditure line shifts vertically from PE_1 to PE_2 , as shown in Figure 6-9. Given the fact that the actual expenditure line ($PE = Y$) doesn't change, the vertical shift in the planned expenditure line causes equilibrium output to increase from \tilde{Y}_1 to \tilde{Y}_2 .



Keynes suggested using government policy to increase planned expenditures by whatever amount was necessary to increase the economy's short-run, sticky-price equilibrium, \tilde{Y} , all the way to the full-employment output level, Y^* . (Famously, he extended this idea to paying some people to dig holes in the ground and others to fill them in again – in other words, any action is better than no action!) In Chapter 7, we discuss such policies in greater detail, including why they don't always work so well in practice.

Putting it all together: The economics of the credit crunch

Explaining the credit crunch is like explaining the plot of *Murder on The Orient Express* - spoiler warning! - they all did it. To get to the events which sparked the banking crisis, you have to understand the base conditions, and those conditions have roots that go back quite a long way. So, in very short and necessarily

brief order, we'll lay them out. In the first place the 1990s boom led to a large quantity of capital being accumulated by lending institutions. The holders of that capital naturally looked for new ways to make the capital work for them, by finding new markets in which to invest. At the same time, Governments, led by the 'Anglo Saxon'

economies including the US and UK, tried to find new ways to encourage people to buy their own homes. (After all, people who own their own home are able to borrow against that asset to fund new businesses, and that, in general, is a good thing for job creation.) However, extending home ownership to those currently not buying homes meant finding a way to manage the risk of poorer debtors defaulting. The solution, to group together bad risks with good ones and sell those as a package, meant that more capital could be lent to people who would otherwise be too risky to lend to. As a result, mortgage lending went up (in part fuelling house price booms) and more people were able to borrow more money to buy their own homes.

This was all fine and dandy while the market kept growing. However, rises in interest rates in the US led to growing defaults, as marginal, higher risk customers became less able to afford their borrowing. As a result, more institutions were left holding more risky debts. As long as banks were able to borrow money from

each other this wasn't a problem. But banks started to figure out that some of the borrowers were holding bad debts, and that they were therefore risky lending options. Before you could say 'lemon market', lending banks began to worry about the bad borrowers out there, and the market for bank lending dried up, just as Akerlof's model (that we discuss in Chapter 15) suggests.

Considering the effect of bank lending on the money supply, this meant that it was harder not only for would-be home-buyers to get access to capital, but also that businesses couldn't get easy access to capital to invest or expand. The knock on effects on the economy, through consumer borrowing, also led to falls in aggregate demand that we call a recession. As of 2010, the consensus seems to be that the response to the crunch, injecting government spending, has had some stabilising effect but it has been expensive and will need to be recouped. Currently the argument is about how fast it is desirable to do this.

Chapter 7

Fighting Recessions with Monetary and Fiscal Policy

In This Chapter

- ▶ Using monetary and fiscal policy to stimulate the economy
 - ▶ Facing the fact that too much stimulus causes inflation
 - ▶ Realising that rational expectations can frustrate monetary and fiscal policy
 - ▶ Getting the details behind monetary and fiscal policy
-

Monetary and fiscal policy are two of the most important functions of modern governments. *Monetary policy* focuses on increasing or decreasing the money supply in order to stimulate the economy, whereas *fiscal policy* uses government spending and the tax code to stimulate the economy.

Thanks to the development of good economic theory, governments these days have a fair idea of how to use monetary and fiscal policy to mitigate the duration and severity of recessions. This development is hugely important because it gives governments the chance to make a positive difference in the lives of billions of people. Good economic policy can make a nation prosperous, whereas bad economic policy can ruin it.

Monetary and fiscal policy are not without problems, however, and in this chapter we show you not only how well they can work under the best-case scenario, but also their limits and problems when implemented in the real world. By seeing the whole picture, you can decide for yourself when and how monetary and fiscal policy should be used.

The information in this chapter puts you two steps ahead of many politicians and helps you to judge when politically biased economists are trying to pull a fast one. As Joan Robinson, one of the great economists of the 20th century, said, 'The purpose of studying economics is not to acquire a set of ready-made answers to economic questions, but to learn how to avoid being deceived by an economist.' We totally agree. And you can trust us, because after all, we're economists.

If you haven't read Chapter 6, we encourage you to do so before tackling this chapter. Although our goal with this book is to have each chapter stand alone so that you can jump in and jump out wherever you need, much of the terminology you encounter in this chapter is introduced and explained in Chapter 6. You may find it easier to tackle monetary and fiscal policy if you have a basic understanding of how recessions work, which is the focus of Chapter 6.

Stimulating Demand to End Recessions

We're going to look at monetary and fiscal policy separately and in detail. First, though, we're going to make an assumption about what each of them is for – altering the aggregate demand for goods and services. (The *aggregate demand* is the total demand for goods and services in an economy.) In particular, both policies can be used to increase aggregate demand during a recession.

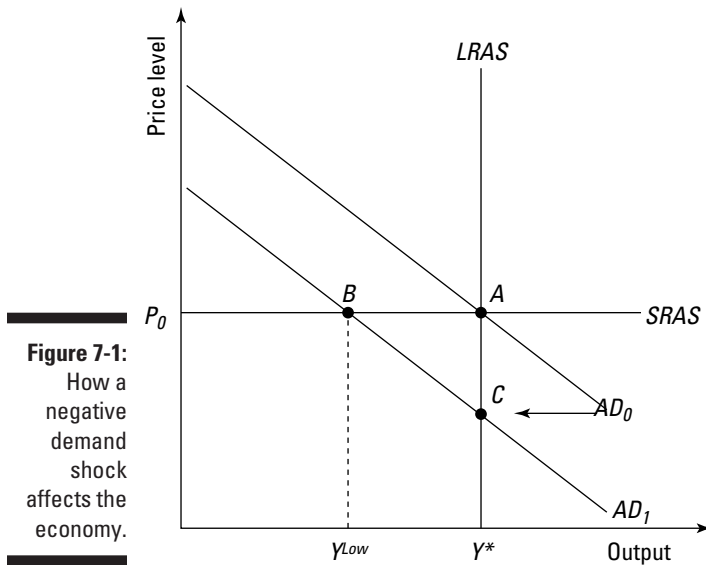
Aiming for full-employment output

The ability to use monetary and fiscal policy to stimulate the economy is important because you always want to end a recession and return the economy to producing at the full-employment output level as quickly as possible.

As we explain in Chapter 6, the full-employment output level – symbolised by Y^* – is the amount of output the economy produces at full employment, which occurs when every person who wants a full-time job can get one. If the economy goes into recession and produces less than Y^* worth of output, millions of people lose their jobs because firms need fewer workers to produce the smaller amount of output.

Worse yet, the unemployment rate remains high until output returns to the full-employment level. Monetary and fiscal policy are useful precisely because they can help return the economy to producing at Y^* as soon as possible; they can shorten the period of frustration and misery that the unemployed endure.

Take a look at Figure 7-1, which shows how monetary and fiscal policy can be used to stimulate aggregate demand and return an economy to producing at Y^* as quickly as possible after the economy is hit with a negative demand shock. (As we explain in Chapter 6, *negative demand shocks* are things that unexpectedly decrease aggregate demand, such as a drop in consumer confidence.)



In Figure 7-1, the economy begins in equilibrium at point A , where the downward-sloping aggregate demand curve, AD_0 , intersects the vertical long-run aggregate supply curve, $LRAS$. As we explain in Chapter 6, prices in the economy are fixed in the short run. For this reason, the short-run aggregate supply curve, $SRAS$, is horizontal at the initial price level (P_0), which is determined by the intersection of AD_0 and $LRAS$. (We explain in Chapter 6 in the section ‘Dealing with fixed prices in the short run’ that for simplicity’s sake we use horizontal $SRAS$ curves in this book instead of the upward-sloping curves used in some other books. If you’re used to seeing upward-sloping $SRAS$ curves, take a quick peek at that section.)

When the negative demand shock comes along, here’s what happens:

- ✓ The aggregate demand curve shifts left to AD_1 , reflecting the reduction in spending on goods and services.
- ✓ With prices fixed at P_0 in the short run, the economy’s equilibrium shifts leftward from point A to point B , and output in the economy falls from Y^* down to Y^{Low} .
- ✓ As output falls, unemployment rises because firms don’t need as many workers.

As you can see, the overall result of the demand shock is a recession: a period of declining output and increasing unemployment.

Unfortunately, a recession can take a long time to resolve. As we explain in Chapter 6, if the government takes no action to end a recession, the only way for the economy to return to producing at the full-employment output level is for prices to drop so that the economy's equilibrium can slide down the AD_1 curve from point B to point C . That process is typically very slow because of sticky prices, and especially sticky wages, which we describe in Chapter 6. As a result, the economy has high unemployment and takes a long time to get back to producing at Y^* unless the government gets involved.

Shifting the AD curve to the right – or, putting people back to work



Monetary and fiscal policy accomplish the trick of increasing aggregate demand, which eliminates the need to endure the slow adjustment process that takes the economy from point B to point C (see Figure 7-1). The policies achieve this aim by shifting the aggregate demand curve to the right.

For example, if the government was able to shift the aggregate demand curve from AD_1 back to AD_0 , the economy would jump back to equilibrium point A . That's very nice because it gets the economy back to producing at Y^* without having to go through the slow adjustment process needed to get an economy to move from B to C . In human terms, the result is that unemployment ends much sooner for millions of workers who can once again find jobs and provide for themselves and their families.

Unfortunately, however, actually implementing aggregate demand shifts to fight recessions isn't easy. Several problems can creep up involving inflation and people's expectations about how increases in aggregate demand affect prices. So before we tackle the details about how monetary and fiscal policy can be used to increase aggregate demand, we first want to explain how inflation (and worries about inflation) can limit their effectiveness.

Generating Inflation: The Risk of Too Much Stimulation

The best way to begin to understand the limitations of economic policies that stimulate aggregate demand is to understand that in the long run, such policies can change only the price level, not the level of output. Why? We need several pages to explain the reasons fully, but the explanation comes back to something we discuss at length in Chapter 6: no matter where the aggregate demand curve happens to be – no matter how much stuff consumers are willing (or unwilling) to buy – prices eventually adjust until the economy is again

producing at full-employment output (Y^*). The economy simply doesn't want to stray from Y^* for too long.

We explain the economy's affection for Y^* in Chapter 6, and you can see it in Figure 7-1 as well. The negative demand shock shifts the aggregate demand curve from AD_0 to AD_1 . If the government doesn't use some sort of stimulus, the economy slowly adjusts on its own from point A to point B to point C . At point C , the price level has fallen and output has returned to Y^* .

But even if the government applies some sort of stimulus to move the aggregate demand curve to the right of AD_1 , the long-run result is always that the economy comes to equilibrium at the point where the aggregate demand curve intersects the long-run aggregate supply ($LRAS$) curve. And, as we show in Chapter 6, the $LRAS$ is a vertical line that corresponds to the full-employment output level, Y^* .

*An exercise in futility: Trying to increase output beyond Y^**

Because the economy always returns to producing at full-employment output (Y^*), the government can't keep the economy producing more output than Y^* for any significant period of time. To see why, suppose that the government uses monetary and/or fiscal policy to shift the aggregate demand curve from AD_0 to AD_1 , as shown in Figure 7-2.

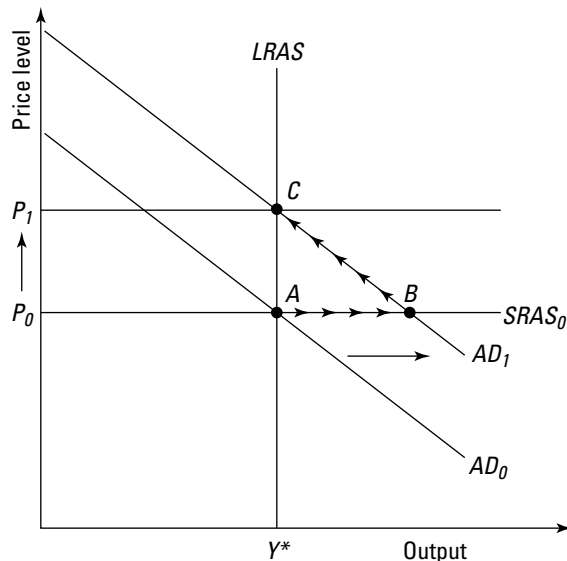


Figure 7-2:
The results of increasing aggregate demand.

Before the shift, the economy is in equilibrium at point *A*, where the original aggregate demand curve, AD_o , intersects the long-run aggregate supply curve ($LRAS$), which is a vertical line above Y^* . At that initial equilibrium, the price level is P_o , and because prices are sticky in the short run (see Chapter 6), the short-run aggregate supply curve, $SRAS_o$, is a horizontal line at P_o .

When the government stimulates the economy and shifts the aggregate demand curve to the right from AD_o to AD_1 , the economy initially shifts from point *A* to point *B*. That is, because prices are fixed in the short run, the economy adjusts to a temporary equilibrium at *B* (where AD_1 intersects $SRAS_o$).

The economy's output level at point *B* is greater than the full-employment output level, Y^* . For this reason, point *B* is only a temporary equilibrium. That's because the economy can only produce more than Y^* by using more labour than is used at Y^* . In our simple model, only two ways exist in which this situation can happen:

- ✔ Firms convince existing workers to work overtime.
- ✔ Firms increase the total number of workers by tempting people such as retirees, who wouldn't normally be in the labour force, to take jobs.

Both ways of increasing the labour supply increase labour costs:

- ✔ To get existing workers to work overtime consistently, firms must pay them high overtime wages.
- ✔ To tempt people such as retirees to join the workforce, firms must increase wages (because, obviously, these people aren't tempted to work at the old wages).

Either way, production costs rise. And as they do, firms pass these costs on to consumers by raising the prices they charge for goods and services.

That's why the economy moves from point *B* to point *C* in Figure 7-2. Prices rise because wages are increasing, and therefore the economy moves up the AD_1 curve (as indicated by the arrows). Wages, and hence prices, continue to rise until the economy is again producing Y^* worth of output at point *C*. At that point, no need exists for further wage increases; the economy is once more producing at Y^* , and firms don't have to increase wages to try to produce more than that level.

A temporary high: Tracing the movement of real wages

If you look at Figure 7-2 and consider the movement from *A* to *B* to *C* caused by the government's stimulus programme, you can see that the only long-run

consequence is an increase in the price level from P_o to P_f . After a period of increased production, the economy is back to producing at the full-employment output level, Y^* .

You can take two critical lessons away from this example:

- ✓ The government can't permanently keep output above Y^* .
- ✓ The government can't permanently keep more people employed than the number employed at Y^* .



These two lessons are true because of real wages. *Real wages* are wages measured not in terms of money, but in terms of how much stuff workers can buy with the money they're paid.

Real wages are crucial to understanding how government stimulus affects the economy, because people don't work hard for money in and of itself – they work hard for the things that money can buy. This distinction is important because as the economy reacts to the government's shifting of the aggregate demand curve from AD_o to AD_f , real wages increase only temporarily. When real wages are higher, workers supply more labour. But when real wages fall back down to their original levels, workers go back to supplying their original amount of labour.

Raising nominal wages while prices are stuck



Confused? Stick with us. To see how the concept of real wages works, consider the situation of a banana-loving worker named Ralph. When the economy is at point *A* in Figure 7-2, Ralph is paid £10 per hour, and his favourite food, bananas, costs £1 per kilogram. This situation implies that his real wages – his wages measured in terms of what they can buy – are 10 kilograms of bananas per hour. At that real wage, Ralph is willing to work full-time.

When the government stimulates the economy and shifts the aggregate demand curve from AD_o to AD_f , workers like Ralph benefit at first because real wages initially rise. That's because in order to produce more output than Y^* , firms have to raise *nominal wages* (wages measured in money) in order to get workers to produce more. Because prices are initially sticky at price level P_o , the increase in nominal wages means an increase in real wages.

In Ralph's case, suppose that the price of bananas remains at £1 per kilogram because of sticky prices, but Ralph's nominal wage rises to £12 per hour because the company he works for needs more labour. Ralph's real wage increases from 10 kilograms of bananas per hour to 12 kilograms of bananas per hour.

This increase in real wages motivates workers to supply all the extra labour that's required to produce higher levels of output. (In Figure 7-2, this event is happening at point *B*.) Because nominal wages have gone up but prices

haven't, the resulting increase in real wages causes workers to supply more labour, which in turn allows firms to produce an output level greater than Y^* .

Moving back to Y^ and to original real wages*

Unfortunately, as firms begin to pass on the costs of increased wages as higher prices, real wages begin to fall. Suppose that because of higher labour costs, the price of bananas rises to £1.10 per kilogram. At that price, Ralph's real wage falls from 12 kilograms of bananas per hour down to 10.91 kilograms of bananas per hour. (To get 10.91, divide Ralph's £12 per hour money wage by the £1.10 per kilogram price of bananas.)

In Figure 7-2, the decrease in real wages happens as the economy moves along the aggregate demand curve from point B to point C . As prices rise, real wages fall. Prices are going to continue to rise until they reach the point where real wages return to where they originally were at point A before the government stimulated aggregate demand.

In Ralph's case, the price of bananas continues to rise until they cost £1.20 per kilogram. At that price, his higher nominal wage of £12 per hour again buys him 10 kilograms of bananas per hour; his real wage is back where it started.



This boomerang effect in the real wage makes total sense. Because the economy returns to producing at Y^* , you only need to motivate workers to supply enough labour to produce Y^* , not anything extra. Workers like Ralph were willing to supply that amount of labour at point A for a real wage of 10 kilograms of bananas per hour. After the economy moves to point C , they're once more willing to supply that amount of labour for the same real wage.

Obviously, not every worker is fixated on bananas like Ralph. But we hope you get the idea: if both wages and prices rise by 20 per cent, real wages remain unchanged and, consequently, the amount of labour that workers supply ends up unchanged.

This fact means that government stimulus policies, such as the one shown in Figure 7-2, which shift aggregate demand from AD_0 to AD_1 , can't permanently increase the amount of labour being employed by firms. Also, these policies can't permanently increase workers' real wages. These effects are at best temporary; they last only as long as the economy takes to adjust from A to B to C .

You may think that a temporary increase in employment and output is pretty good, however, and that the government should still go ahead and increase aggregate demand from AD_0 to AD_1 . Unfortunately, as we're about to show you, if people know about the stimulus ahead of time, the economy may adjust directly from A to C and eliminate the ability of the aggregate demand shift to stimulate the economy even temporarily.

Failing to stimulate: What happens when a stimulus is expected

In the previous section, we explain why an increase in aggregate demand that tries to increase output beyond Y^* can do so only temporarily, until prices adjust. In this section, we show you that prices may adjust so quickly that the stimulus may fail to increase output at all, even temporarily.

Respecting the importance of price stickiness

As we show in Figure 7-2, any increase in output after aggregate demand shifts rightward from AD_0 to AD_1 depends on prices being sticky in the short run. In other words, the economy moves from point A to point B along the horizontal short-run aggregate supply curve, $SRAS_0$, only if the price level is fixed at P_0 in the short run.

In Chapter 6, we explain that a lot of evidence shows that prices have a hard time falling during a recession. In particular, firms don't like to cut wages and insult their workers. They know that if they cut wages, workers become angry and refuse to work as hard, and the resulting decline in productivity makes the firm's profit situation even worse.

As a result, a lot of *downward wage stickiness* is present in the economy – by which economists mean that nominal wages decline only rarely. As we explain in Chapter 6, downward wage stickiness leads to *downward price stickiness*, because firms can't cut their prices below production costs if they want to turn a profit and stay in business. (Keep in mind that labour costs are, for most businesses, the largest part of production costs. If firms can't cut wages, they can't cut the price of their output.)

Realising that prices aren't very sticky upward

Notice that in the previous section we talk only about *downward* stickiness; we don't say anything about prices or wages having trouble rising. In fact, very little in the economy seems to cause *upward* wage stickiness or *upward* price stickiness.

Quite the contrary, wages and prices seem quite free to rise if demand increases relative to supply. Business contracts and labour contracts may limit price and wage increases for a while, but as soon as these contracts expire, prices and wages are free to rise.

Anticipating (and undermining) a stimulus

The lack of upward price stickiness implies two very important things for any government attempting to stimulate the economy into producing more than the full-employment output level (Y^*):

- ✓ If prices and wages can rise quickly, the economy produces more than Y^* only very briefly. That is, it moves from A to B to C in Figure 7-2 very quickly – so quickly that the stimulus causes output and employment to rise above Y^* only very briefly.
- ✓ If people can see a stimulus coming, that stimulus (which attempts to increase output beyond Y^*) is likely to generate only inflation and no increase in output whatsoever. In other words, if people can anticipate an increase in aggregate demand, the economy may jump directly from point A to point C , so that the price level rises without even a temporary increase in output.



To understand how this process works, suppose that the government preannounces a big stimulus package that's going to shift aggregate demand from AD_0 to AD_1 in a few months' time (see Figure 7-2). Because workers and businesses can find out about macroeconomics just as well as the politicians running the government, they realise that the only long-run effect of the upcoming stimulus will be for prices to rise from P_0 to P_1 .

In addition, workers understand that real wages will remain unchanged in the long run, because both their nominal wages and their cost of living (given by the price level) will increase by equal amounts. As a result, they know that in the long run, the stimulus isn't going to help them at all. Indeed, their only hope for gains is based entirely upon the short run, when nominal wages should go up and the price level should stay the same. In other words, they hope to benefit from the movement from A to B in Figure 7-2.

But firms aren't stupid. They don't want to have their profits reduced because wages are rising while prices are fixed. So they simply anticipate everything. Because prices eventually have to rise from P_0 to P_1 , and wages eventually have to rise by an equal amount, firms get ahead of the wage increases by raising prices as soon as they can.

Nothing prevents firms from raising prices because nothing in the economy is causing upward price stickiness. So, if firms can see the stimulus coming ahead of time, they simply raise prices as soon as they can in order to make sure that prices and wages are going up at the same pace. As a result, the price level jumps from P_0 to P_1 .

Of course, at the same time, firms raise wages by an equal percentage in order to keep real wages the same. They want to keep workers motivated to supply the labour necessary to produce Y^* worth of output.

As you can see, if a government tries to stimulate the economy to produce beyond Y^* , and if the stimulus is understood and anticipated by everyone in the economy, the policy may not work at all. Prices and wages may simply

jump from point A to point C , meaning that the stimulus fails to stimulate because output stays constant at Y^* while prices and wages go up simultaneously.

Having rational expectations

The phenomenon we describe in the previous section is an example of *rational expectations*, a term that economists use to describe how people rationally change their current behaviour in anticipation of future events. In this case, firms rationally decide to raise prices immediately when they find out that the government is going to increase aggregate demand from AD_0 to AD_1 in the future.

Indeed, firms' only rational course of action is to raise prices immediately, because if they leave prices alone at P_0 , they're volunteering for the decrease in profits that results when the economy moves from point A to point B (when nominal wages rise while prices stay constant). By immediately raising prices and shifting the economy directly from A to C , firms can avoid that situation altogether.



Rational expectations is one of the most important ideas in macroeconomics because it tells you that strong limits constrain the government's ability to control the economy. People don't just sit around like potted plants when the government announces a policy change. They change their behaviour. And sometimes, as in the case we describe in the previous section, their behavioural change completely ruins the government's ability to achieve its objective of stimulating the economy.

As we explain monetary and fiscal policy in more detail in the rest of the chapter, notice the other examples of rational expectations limiting the effectiveness of government policy. In particular, notice how, in every case, changes in people's behaviour reduce the impact of government policy initiatives.

Figuring Out Fiscal Policy

Fiscal policy concerns itself with how governments tax and spend. This policy overlaps with macroeconomics because modern governments have many opportunities to increase aggregate demand by making changes in fiscal policy. These changes fall into two main categories:

- ✓ Increasing aggregate demand indirectly by lowering taxes so that consumers have larger after-tax incomes to spend on buying more goods and services
- ✓ Increasing aggregate demand directly by buying more goods and services

The first category involves decreasing government revenues, and the second involves increasing government spending. Because the government's budget deficit is defined as tax revenues minus spending, both types of fiscal policy are likely to increase government budget deficits. This fact is very important because large and ongoing government budget deficits may lead to many economic problems, including inflation. As a result, the fear of large budget deficits constrains the magnitude of fiscal policy initiatives.

As you read about fiscal policy in the following sections, keep this fear of large budget deficits in mind because it limits the size of the aggregate demand shifts that a government can undertake. For example, if you look back at Figure 7-1, the government may want to use fiscal policy to shift aggregate demand rightward from AD_0 to AD_1 , but if doing so involves an overly large budget deficit, the government may have to settle for a smaller shift that moves the economy only part of the way back to producing again at full-employment output (Y^*).

Increasing government spending to help end recessions

If an economy gets into trouble, one of the options open to policymakers is to increase government spending. The idea is that if people are unemployed and unsold goods are sitting around gathering dust, the government can come in with a lot of money and buy up a lot of the unsold products. The result of this action is that the government generates so much demand that businesses start hiring the unemployed in order to increase output to meet all the new demand.

The hope is that this stimulus jumpstarts further demand. When people who were formerly unemployed start getting paid again, they start spending more money, which means that demand rises. When this happens, the economic recovery should be self-sustaining so that the government doesn't need to continue to spend so much money.

Paying for increased government spending

Politicians naturally like suggesting increases in government spending because such increases make them look good, especially if they can get some of the new spending earmarked specifically for their own constituents. However, nothing in life is free.

Only three ways exist to pay for increased government spending:

- ✓ The government can lower interest rates to expand the money supply.
- ✓ The government can raise taxes.
- ✓ The government can borrow more money.



As we discuss in Chapter 5, expanding the money supply to pay for increased government spending leads to large bouts of inflation, which brings economic chaos and recessions. Consequently, governments nowadays almost never resort to doing so. In fact, many have given up control of money supply to independent central banks. Although, as we note in Chapter 5, you could use this tool to get out of recessions, it's not really done now.

Raising taxes is also problematic, because if you're trying to get out of a recession you want consumers to spend as much as possible on goods and services. If you raise taxes, consumers reduce their spending. You may offset some of the decreased private spending by immediately turning around and spending all the tax revenue, but clearly this approach is no way to stimulate aggregate demand in the long run. The government may as well just let its citizens spend their money in the first place.

Borrowing and spending: The most common solution

What governments need to do to combat recessions is figure out a way to increase their own spending without decreasing private spending. The solution is borrowing.

By borrowing and then spending money during a recession, the government can increase its purchases of goods and services without decreasing the private sector's purchases. Who does the government borrow from? You, and other people like you.

At any given moment in time, people want to save a certain part of their incomes. They can use these savings to buy many different kinds of assets, including stocks and bonds issued by corporations, real estate, mutual funds and annuities. But they can also use their savings to buy government bonds, which are, in essence, loans to the government.



By offering more bonds for sale, the government can redirect some of the savings that people are making away from purchases of other assets and into purchases of government-issued bonds. By selling bonds, the government can get hold of lots of money to spend on goods and services, thereby turning what would have been private spending on assets into public spending on goods and services.

Dealing with deficits

Increasing government spending and financing it through borrowing is clearly a good way to increase the overall demand for goods and services. But it has the potentially nasty side effect of creating a *budget deficit*, which is the amount by which government spending exceeds tax revenues during the current year. Any current budget deficit adds to the *national debt*, the cumulative total of all the money that the government owes lenders.

The problem with budget deficits and the national debt is that they have to be paid back someday. Consider a ten-year bond that pays a 6 per cent rate of return. When you buy the bond from the government, you give it, for example, £1,000. In return, the government promises to do two things:

- ✔ To give you back your £1,000 in ten years
- ✔ To give you £60 per year (a 6 per cent return) until you get your £1,000 back

So, the government gets £1,000 right now to spend on goods and services to boost the economy, but it needs to figure out where to get £60 per year to provide your interest payments, and also where to get £1,000 in ten years when the bond matures.

Relying on the security of future tax revenues

Obviously, people are only willing to lend the government money by buying bonds because they believe that the government is eventually going to pay them back. They have confidence in that happening because governments have the exclusive right to tax things. Essentially, future tax revenues secure all government borrowing.



But the link between taxes and bond repayments isn't direct. In other words, just because a government has a lot of bonds coming due doesn't necessarily mean it has to raise taxes suddenly to get the money to pay off the bonds. Instead, governments often refinance the bonds that are coming due; they simply issue new bonds to get enough cash to pay off the old bonds. This process is referred to as *rolling over the debt* and is routinely practised by governments everywhere.

But don't think that this process is a huge scam to defer paying off the debt indefinitely. The only reason that investors are willing to participate in a roll-over is that they have confidence that the government can always use its tax powers to pay off its debts. Investor confidence allows governments to keep on borrowing, whether to fund new borrowing or to roll over old debt. When investors don't have that confidence, the effect can be catastrophic.

Paying the debt by printing money: A devastating choice

Sometimes, investor confidence in the government turns out to have been misplaced. As we discuss in Chapter 5, governments have another (rather diabolical) way to pay off their bonds besides using tax revenues: they can expand the money supply. Responsible governments generally don't take this action any longer, but it has happened, and it's really worth being aware of the possibility.



A £1,000 bond obligates the government to pay you back £1,000 worth of money. The bond doesn't say where that £1,000 comes from. So the government is free to print £1,000 worth of new bills and hand them to you. This solution may seem okay at first, but when you and all the other bond holders with newly printed cash go out into the economy and start spending that new money, you drive up prices and cause inflation.

As we point out in Chapter 5, high inflation destroys economic activity. During high inflation, prices lose much of their meaning, and people are much more mistrustful and reluctant to engage in long-term contracts or make long-term investments because they don't know how much money will be worth in the future.

Knowing the potential horrors of inflation, people tend to worry any time they see a government running large budget deficits or piling up a very large debt. They worry that the government may find itself in a position in which it can't raise taxes high enough to pay off its obligations (or is unwilling to anger voters by raising taxes that high). Investors worry that if this situation occurs, the government may resort to printing money to pay off its debts. And doing so ruins the economy.

Printing money to pay government debts also badly hurts most bondholders because most of them get their cash after prices have gone up, meaning that their cash doesn't buy much stuff. Consequently, when people really begin to worry that a government may start printing money to pay off its debts, the government finds getting anyone to buy its bonds harder and harder. In such a situation, to get anyone to buy its bonds, the government offers higher and higher interest rates to compensate for people's worries that the money they're eventually going to get back isn't going to be worth much. These higher interest rates then make the government's situation even more desperate, because any debt rollovers have to be done at the higher interest rates.

Furthermore, because inflation affects all bonds, not just the ones issued by the government, interest rates all across the economy rise if people fear inflation is coming. This situation can have bad economic consequences *immediately*, because higher interest rates dissuade consumers from borrowing money to buy things like cars and houses, and also discourage firms from borrowing money to buy new factories and equipment. Consequently, just the expectation that a government may print money at some point in the future to pay off its bonds can cause immediate harm to the economy. (This example again shows rational expectations in action; see the section 'Having rational expectations' earlier in the chapter.)

Most governments try to keep their debt level and their deficits under control to ensure that no one seriously worries that the government is ever going to be tempted to print money to pay off its bonds.

Dissecting Monetary Policy



Monetary policy is the manipulation of the money supply and interest rates in order to stabilise or stimulate the economy. In modern economies, monetary policy has come to be regarded as the most powerful mechanism that governments have at their disposal to fight recessions and reduce unemployment – even more powerful than fiscal policy.

Monetary policy is put into practice by first changing the supply of money in order to manipulate interest rates. Because interest rates affect everything from the demand for home mortgages by consumers to the demand for investment goods by businesses, they have a huge and pervasive effect on stimulating or depressing economic activity.

To give you a complete picture of how monetary policy functions, we first review what money is. We then show you that an economy can have too much money, and that this fact is related to interest rates and inflation. In turn, this information gives you the necessary insight to understand how the government can affect interest rates by changing the amount of money that's floating around in the economy.

Identifying the benefits of fiat money

Money is an *asset*, meaning that it holds its value over time. Other assets include real estate and property, precious metals like gold and financial assets like stocks and bonds. But money is usually held to be a better medium of exchange than any of the above.

To keep things simple, we've referred to the concept of printing money a lot. The thing to remember, though, is that 'money' does not just mean the cash and coins in circulation, but has a wider definition that includes cash deposits in banks, credit and loans, and government bonds of varying degrees of tradeability. As a result, we also talk about *expanding the money supply*, which means doing something that increases the supply of all of the above.

As we explain in Chapter 5, money makes an economy much more efficient because it eliminates the need to engage in barter. But the need to verify the authenticity of money (so that people are willing to accept it) means that the responsibility for producing money and suppressing counterfeits has fallen to governments.

This situation, in turn, brings its own potential problems, because governments always face the temptation to print more money in order to pay off old debts or buy lots of newly produced goods and services.



Historically, one way to limit governments' ability to print up more money to pay off bills was to put them on a *metallic standard*. Under such a system, governments were unable to print more bills without backing them with a precious metal, like gold. For example, the United States used to have a gold standard under which \$35 of currency was redeemable for one ounce of gold. You were literally able to bring \$35 of bills to the US Treasury and exchange the cash for an ounce of gold.

For monetary policy, this situation meant that the government was unable to increase the supply of paper money arbitrarily because for every \$35 of new bills it wanted to print, it had to buy an ounce of gold with which to back them. The high cost of buying gold limited the money supply.

As we note in Chapter 5, such a system is great for preventing high inflation because the only way you ever get high inflation is if the government prints a huge amount of new money. (When that new money begins circulating, it drives up prices.)

Preventing inflation is a good thing, but using a metallic standard turns out to have some large drawbacks. Using a metallic standard causes the supply of money to be pretty much fixed over time, meaning that even if the economy needs a little bit more or a little bit less money to make it work better, the government can't do anything because the supply of money is fixed by the amount of gold the government has in its vaults.



In particular, the metallic standard means that you can't use monetary policy to stimulate your economy if it gets into a recession. One of the reasons that the Great Depression was so bad everywhere around the world was that nearly every country was on a gold standard when the calamity began. This arrangement meant that governments were unable to increase their money supplies in order to help their economies. This situation also explains why the countries that quit their gold standards earliest had the shortest and mildest recessions; after they quit, they were free to print new money to stimulate their economies. On the other hand, countries like the United States and the United Kingdom that stubbornly stuck to their gold standards had the most prolonged and painful economic downturns.

This was obviously a problem. The attempt to keep to a metallic standard broke down later as the US government found it needed to create more money to pay for the escalating costs of the Vietnam War. This led to the adoption of a system called *fiat money*. Under a fiat money system, the government simply prints up as many bills as it likes, declares them to be money, and puts them out in the economy. (*Fiat* means 'let it be' in Latin.) The great benefit of this system is that the government can arbitrarily increase or decrease the money supply in whatever way best helps to stimulate the economy.

For the rest of this chapter, we use M to denote the total supply of money floating around the economy. For example, $M = \text{£}1.3$ trillion means that the sum of the face values of all the bills and coins in the economy is $\text{£}1.3$ trillion.

Realising that you can have too much money!



Monetary policy works by manipulating the supply of money in order to change the price of borrowing money, which is the interest rate. The key to making monetary policy work is the fact that the demand for money depends on the interest rate.

Imagine that we hand you $\text{£}1$ million, and you can do whatever you want with it. Suppose that you're frugal and decide to save every last penny, at least for a year, because you think that's going to give you enough time to figure out how to best blow the money.

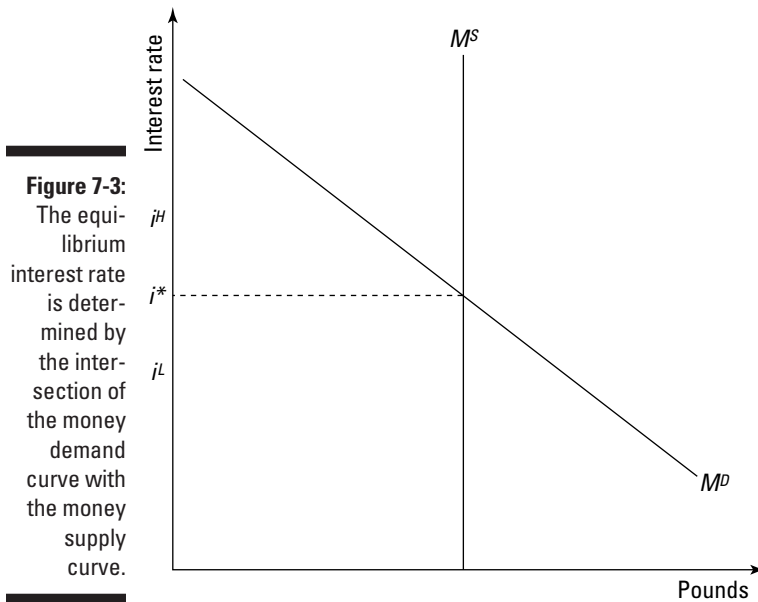
Our question to you is: should you keep all your new wealth in cash?

The correct answer is 'NO!'.

Holding your wealth in cash is, to be blunt, really stupid because cash earns no interest. Even if you put the money into a standard current account, you get at least a tiny bit of interest. Even 1 per cent of interest on a million pounds is $\text{£}10,000$. Why would you give that up? Even better, if you use the cash to buy government bonds, you may get 5 or 6 per cent. That's $\text{£}50,000$ or $\text{£}60,000$ more than if you keep your wealth in the form of cash.

Clearly, the higher the interest rate you can get on other assets, the more incentive you have to convert your cash into other assets. In fact, the only thing preventing people from converting all their wealth to other assets and never holding any cash is the fact that money lets them buy things. Beyond that function, money is no better than any other asset; in fact, it's worse in terms of its rate of return because the rate of return on cash is always zero.

In Figure 7-3, we create a graph that demonstrates how much money people demand to hold at any particular interest rate. We denote money demand as M^D . The nominal interest rate, i , is on the vertical axis. (For an explanation of nominal interest rates, see Chapter 5.) The horizontal axis is measured in currency, which in the UK means pounds (until the country adopts the euro, whenever that may be).



As you can see from the downward slope of the money demand curve, the higher the interest rate, the less money people want to hold. This graph simply represents the idea that cash, with its zero interest rate, is a worse and worse place to park your wealth if you can get higher and higher returns in alternative assets. In other words, the higher the interest rate on other assets, the more you're going to want to economise on your cash holdings.

Figure 7-3 also contains the vertical money supply curve, where M^S stands for money supply. This curve is vertical because the government can decide how much money it wants to print and circulate without regard to the interest rate.

The M^D and M^S curves cross at interest rate i^* . This interest rate is the *equilibrium* interest rate, because it's the only rate at which the total amount of money that people want to hold is equal to the total amount of money that the government has circulated.

More importantly, i^* is a *stable equilibrium*, meaning that if interest rates ever deviate from it, market forces are going to push them back to i^* . But before this fact is going to make sense, we need to take a few paragraphs to explain how interest rates are determined in the bond market. Pay close attention because bond markets are *the* place where interest rates for the whole economy are determined. Bond markets have a huge effect on everything else that goes on in the economy.

Getting to know bond . . . government bond

A *bond* is a financial asset for which you pay a certain amount of money right now in exchange for a series of payments in the future. Two kinds of payments exist, face value payments and coupon payments:

- ✓ The *face value payment* is printed on the face of the bond certificate and comes on the date the bond expires.
- ✓ The *coupon payments* are typically made twice per year until the bond expires. They're called *coupon payments* because before computerised recordkeeping, you literally clipped a coupon off the bottom of the bond certificate and mailed it in to receive your payment.

Typically, bonds expire after one, five, ten or twenty years.

Bonds don't guarantee any sort of rate of return. They promise only to make the coupon and face value payments on time. The rate of return depends on how much you pay for the right to receive those payments.



If you think we're speaking in tongues right now, bear with us. Imagine a really simple kind of bond called a *zero-coupon bond* (so named because no coupon payments apply). The only payment this bond ever makes is the face value payment that comes when the bond expires. And to make things really simple, suppose that it pays its owner exactly £100 exactly one year from now.

If you're the bond owner, you have to understand that the rate of return the bond pays depends on how much you pay for it right now. Suppose that you were naive enough to pay £100 for the bond right now. Your rate of return would be zero per cent, because you paid £100 for something that's going to give you £100 in a year.

On the other hand, suppose that you pay only £90 for the bond right now. Your rate of return is going to be about 11 per cent, because $(£100 - £90)/£90 = 0.111$, or 11.1 per cent. If you were able to buy the bond for only £50, your rate of return would be 100 per cent, because you'd double your money in a year's time.

You can work through to the result, but it's really useful to state again here that the rate of return on a bond *varies inversely* with how much you pay for it. Because the amount of money you get in the future is always fixed, the more you pay for it right now, the less your rate of return. Higher bond prices imply lower rates of return.

Seeing the link between bond prices and interest rates

The fact that bond prices vary inversely with interest rates is the key to understanding why i^* is a stable equilibrium in Figure 7-3. In this section, we explain the link.

First, consider interest rates that are higher than i^* , such as i^H . When interest rates are higher than i^* , the amount of money supplied exceeds the amount of money demanded. This situation means that people have been given more of the asset called money than they want to hold. So they try to reallocate their portfolio of assets by using the excess money to buy other assets.

One of the things that people buy is bonds. But with all this new money being thrown at the limited supply of bonds, the price of bonds rises. Now be careful. What happens to interest rates when bond prices rise? They *fall*. That's why if you start out at an interest rate that's higher than i^* , interest rates fall back down toward i^* . Excess money drives up the price of bonds, which lowers interest rates.

On the other hand, for interest rates like i^L that are lower than i^* , the amount of money demanded exceeds the amount of money supplied. Because people want more money than they have, they're going to try to get it by selling non-cash assets like bonds in order to convert those assets into the cash they want.



Imagine that everybody tries to sell his or her bonds to achieve this aim. With all the selling, bond prices fall, meaning that interest rates *rise*. In fact, bond prices continue to fall and interest rates continue to rise until they are back at i^* , because that's the only rate of interest at which people are satisfied holding the amount of money, M^S , that the government has decided to circulate.

You need to understand that the movements back to the equilibrium interest rate, i^* , are very quick. Any excess money demand or excess money supply never lasts very long because rapid adjustments in the price of bonds move the interest rate to its equilibrium.

If interest rates adjust very quickly, it follows that a government could expand the money supply simply as it wishes. This could give the government a useful tool to manage the economy. However, if people know that the government can do this, they simply factor that into their guesses about the future and behave as if the government has already done so. Thus, the usefulness of this policy would be reduced. This is one of the underlying reasons for governments delegating monetary policy to some other authority (such as the Bank of England).

Changing the money supply to change interest rates

Monetary policy works because governments know that interest rates adjust in order to get people to hold whatever amount of money the government decides to print. The interest rate is, in some sense, the price of money, and it reacts in a way similar to other prices. That is, if the money supply suddenly increases, the price of money falls, and vice versa.

You can see this fact graphed in Figure 7-4, in which the government increases the money supply from M^S_0 to M^S_1 . This action shifts the vertical money supply line to the right and lowers the equilibrium nominal interest rate from i^*_0 to i^*_1 .

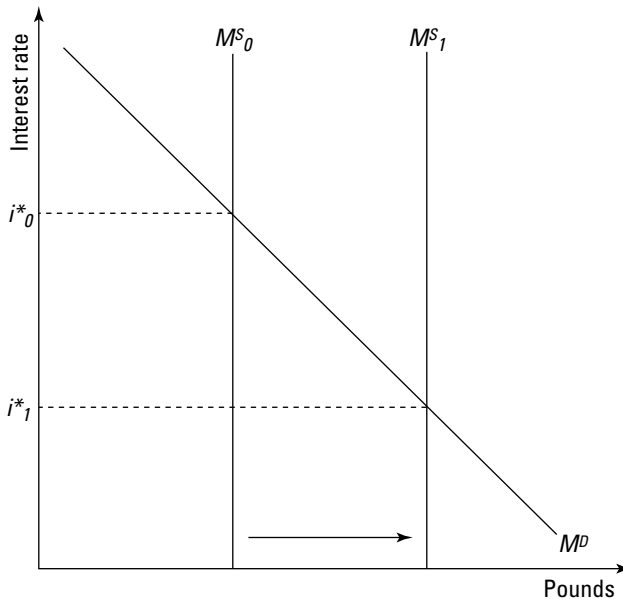


Figure 7-4: Increasing the money supply lowers the equilibrium nominal interest rate.



The system in the UK works as follows. The Bank of England controls the supply of money because it has a *legal monopoly* to do so. Until 1997, it did so at the behest of the Chancellor of the Exchequer, but in a dramatic and widely accepted policy change, the incoming Labour government made the Bank independent of the government. In general, the way the Bank changes the money supply is not by telling the mint to print more or less of the coin of Her Majesty's realm, but by setting something called the *base rate*, which is the lowest possible interest rate that commercial banks can use to borrow from the government – in other words, doing everything we talked about earlier but backwards! Another possibility, and one that's worth discussing, is

using a slightly more subtle method called *open-market operations*. We look at this method now using a notional central Bank, as most of this is internationally standard these days (maybe because central banks are big employers of graduates in macroeconomics).



The term *open-market operations* refers to the central bank's buying and selling of government bonds. That is, open-market operations are transactions that take place in the public, or open, bond market. Depending on whether the Bank buys or sells bonds, the money supply out in circulation in the economy increases or decreases:

- ✓ If the Bank wants to increase the money supply, it buys bonds because in order to buy bonds the government must pay cash, which then circulates throughout the economy.
- ✓ If the Bank wants to decrease the money supply, it sells bonds because the people to whom the Bank is selling the bonds have to give the government money, which reduces the amount in circulation.

By buying or selling bonds in this way, the amount of money out in circulation (M^S) can be very precisely controlled, meaning that the government can, in turn, keep tight control over interest rates.

Lowering interest rates to stimulate the economy

Now that you understand the actual mechanics by which the Bank of England (or similar institutions in other countries) manipulates interest rates, you're ready to see how monetary policy affects the economy.

The basic idea behind monetary policy is that lower interest rates cause both more consumption and more investment, thereby shifting the aggregate demand curve to the right. Here's how:

- ✓ Lower interest rates stimulate consumer consumption spending by making it more attractive to take out loans to buy things such as cars and houses.
- ✓ Lower interest rates stimulate investment spending by businesses because at lower interest rates, a larger number of potential investment projects become profitable. That is, if interest rates are 10 per cent, businesses are only willing to borrow money to invest in projects with rates of return of more than 10 per cent. But if interest rates fall to 5 per cent, all projects with rates of return higher than 5 per cent become viable, and so firms take out more loans and start more projects. (For more on how interest rates affect investment, see Chapter 4.)

When trying to remember how monetary policy works, keep in mind that it's actually a very simple three-step process. If a central Bank wants to help increase output, it initiates the following chain of events:

- 1. The Bank buys government bonds in order to increase the money supply.**
- 2. The increased money supply causes interest rates to fall because the prices of bonds get bid up.**
- 3. Consumers and businesses respond to the lower interest rates by taking out more loans and using the money to buy more goods.**

The hard part is remembering the counterintuitive fact that higher bond prices mean lower interest rates. But if you have a hard time remembering that, don't be embarrassed. Many economists get stuck on it too.

Understanding how rational expectations can limit monetary policy

The government's ability to use increases in the money supply to stimulate the economy is limited by rational expectations and the fears that people have about inflation. Specifically, investors understand that increases in the money supply can cause inflation (as we discuss in Chapter 5). This understanding means that whenever a central Bank increases the money supply in order to lower nominal interest rates, it has to do so with some moderation, in order to avoid causing inflationary fears that can offset the stimulatory effect of increasing the money supply.

Graphing the results of money supply increases

Take a look at Figure 7-5, which shows an economy in recession at point A where aggregate demand curve AD_0 intersects short-run aggregate supply curve $SRAS_0$, which is fixed at price level P_0 . The Bank then increases the money supply to lower interest rates and stimulate the economy, which causes the aggregate demand curve to shift rightward to AD_1 .

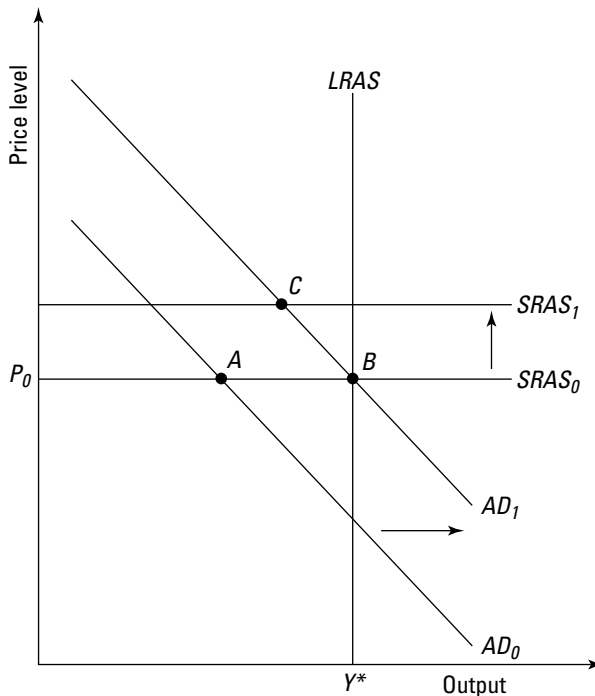


Figure 7-5:
The result of increasing the money supply depends on inflationary expectations.

At this point, two things can happen, depending on people's inflationary expectations:

- ✓ If people believe that the price level is going to remain fixed at P_0 , the rightward shift in aggregate demand moves the economy's equilibrium rightward along the $SRAS_0$ curve from point A to point B .
- ✓ If people believe that the price level is going to jump in response to the increase in the money supply, the short-run aggregate supply curve shifts up vertically by the amount that the price level is expected to increase. Therefore, the economy's equilibrium moves from A to C , where AD_1 intersects the new short-run aggregate supply curve, $SRAS_1$.

Because output increases less if the economy moves from A to C than if it moves from A to B , the Bank obviously has to be careful about inflationary

expectations when trying to stimulate the economy by increasing the money supply. If people expect inflation to occur, their actions can offset some of the stimulus that an increased money supply is expected to bring.

Realising how inflationary expectations affect interest rates

The underlying problem is that the Bank has only partial control over interest rates. In particular, the Bank controls money supply but not money demand. This situation causes a problem: if people think that an increase in the money supply is going to cause inflation, they increase their money demand because they're expecting to need more cash to buy things at higher prices.

So although the increase in the money supply tends to lower interest rates, as shown in Figure 7-4, the increase in money demand caused by inflationary fears tends to increase interest rates. Because higher interest rates tend to decrease investment, any increase in interest rates caused by inflationary fears works against the stimulus that the Bank is attempting to apply to the economy by increasing the money supply.

This decrease in the effectiveness of monetary stimulus is why the big shift in aggregate demand in Figure 7-5 doesn't shift the economy all the way back to producing at Y^* . With people expecting inflation, part of the stimulus ends up causing inflation rather than stimulating the economy to produce more output.

So how does the Bank set monetary policy if it seems to have little or no effect? In fact, two hidden mechanisms are at play.

- ✓ **Precommitment:** By establishing a credible commitment to a monetary policy, the Bank is able to keep everyone else honest. If the Bank says, 'I will act to ensure that inflation is no higher than 5 per cent', you know that it will act to prevent inflation rising any higher. As a result, people will factor the Bank's policy into their decisions and act as if the Bank had already done so.
- ✓ **Advantage:** The Bank knows what it may do when we can only guess, so if it thinks our expectations are out of line with policy, it can pull a surprise move and generate either a short-run stimulus or contraction before we know it.

Keeping inflationary expectations low to help monetary policy work well

Since the 1970s, most countries have been more cautious when using monetary policy. During the 1970s, countries took to heart the lesson of the previous section – that if people believe an increase in the money supply is going to cause inflation, an increase in the money supply may mostly end up causing inflation rather than providing stimulus.

An extreme case of this situation can be seen in Figure 7-6, where output remains unchanged at the recessionary level Y^{Low} , despite an increase in the money supply that causes aggregate demand to shift rightward from AD_0 to AD_1 . The problem is that higher inflationary expectations cause the short-run aggregate supply curve to shift up vertically from $SRAS_0$ to $SRAS_1$, fully offsetting the increase in aggregate demand. The short-run equilibrium shifts from A to B , but the only effect is a higher price level with no increase in output.

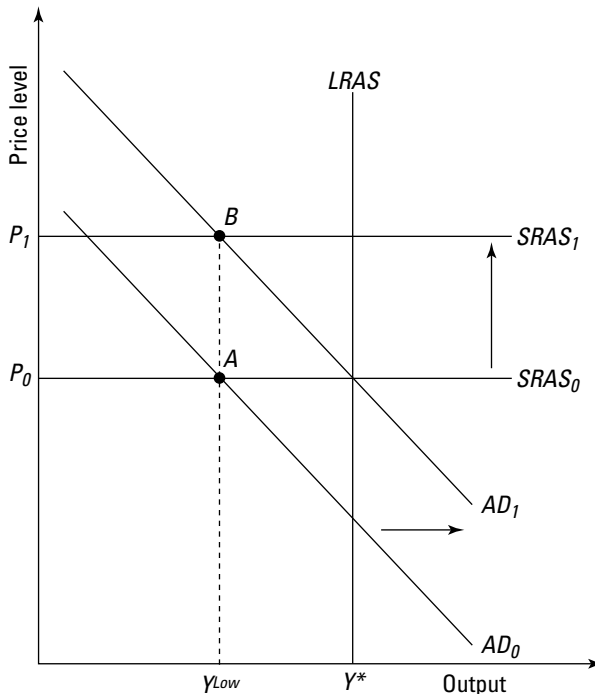


Figure 7-6:
An example
of stagfla-
tion.

The situation in Figure 7-6 came to be referred to as *stagflation*, by which economists meant that the economy simultaneously had a stagnant output level coupled with inflation.



The experience of stagflation during the 1970s taught the Bank of England (and its equivalents in other countries) that monetary policy works best if people believe that the Bank is *not* going to cause inflation. Consequently, these days, the Bank makes only moderate increases in the money supply when it wants to stimulate the economy. These increases end up being more effective than larger increases because they don't trigger inflationary fears. Most central banks these days, especially those in Europe, have *inflation targets* to keep their eyes on too. This situation means that they are required, by

contract with the government or through legislation, to keep inflation within a certain range, and are therefore very circumspect about increasing the money supply. The old German Bundesbank, which forms the model for the European Central Bank, had a reputation for managing the economy soundly in this way.

What do we do now?

Having been presented with the tools policy makers use to control the macroeconomy, you'd be forgiven for believing that the current economic conditions are easily soluble and hardly need further discussion. However, as you'll probably be sick of being told by now, it's not quite so simple.

Crucially, the choices for policy makers are sensitive to many factors. First of all, the cure depends on the diagnosis – and many people disagree on the diagnosis, even when presented with similar symptoms (lower output). Some believe that demand is still deficient, and some that fiscal policy, which might have mitigated the worst effects of the recession, has been overused. The problem, however, is that while data might eventually settle the matter, the reality is that policy makers operate in a climate where wait and see isn't a realistic option. So, policy makers have to decide using what data they have, and that might not give as full a picture as they wish.

The second major issue is that economics isn't like engineering. If you have a machine that goes wrong, you can usually fix matters by tracking down the defective part and replacing it. But an economy is more like a rainforest than a machine; it's a complex web of relationships that can change in unanticipated ways. People change their behaviour based on their best guesses at how the world looks, or is going to look, and economic interventions tend to change the way the world looks.

So crucially, what policy you advocate now depends on a lot of beliefs about the world, both your own and other agents in the model. If you believe on the basis of data that we've overused fiscal policy, you'd prefer to advocate cutbacks. If, however, you believe that we still have spare capacity, or that we're still in danger of sliding back into recession, you'd be likely to advocate further stimulus of the economy. What you're unlikely to be able to do is to decide perfectly between these situations.

Part III

Microeconomics: The Science of Consumer and Firm Behaviour



'Are you hiding something from us, Mr Dingwall?'

In this part . . .

Microeconomics focuses on the decision-making behaviour of individual people and individual firms. In this part, we show you that economic models assume that individuals make decisions in an attempt to maximise happiness, and firms make decisions in an attempt to maximise profits. The pleasant but surprising thing is that in the context of competitive markets, firms pursuing profits and individuals pursuing happiness end up using society's limited pool of resources in the most efficient manner possible – meaning that properly functioning competitive markets produce the best combination of goods and services from society's limited pool of resources. However, markets aren't always set up correctly, so we also cover situations such as monopolies and 'lemons markets' to show you what happens when things go wrong, and how they can be fixed.

Chapter 8

Supply and Demand Made Easy

In This Chapter

- ▶ Explaining why a higher price decreases the quantity demanded of a good or service
 - ▶ Showing why a higher price increases the quantity supplied of a good or service
 - ▶ Demonstrating that demand curves slope down whereas supply curves slope up
 - ▶ Focusing on market equilibrium
 - ▶ Understanding how shifts in demand or supply affect market equilibrium
 - ▶ Identifying policies that prevent market equilibrium
-

Modern economies continually trade goods for other goods, either directly or via the medium of money. All this activity can get very complicated, so we need a simple model to help make sense of it all. This chapter introduces you to just such a simple model, called the *supply and demand model*, to help you make sense of how the quantity of a good sold is related to the price of the good.

For example, the supply and demand model can tell you why the price of hotel rooms goes up during the summer and why the price of wheat goes down after a good harvest. The model can also predict – correctly – that intervention to support the price that farmers receive for crops causes an overproduction of food and that rent controls lead to a shortage of housing.

If you have time to find out about only one aspect of economics, make sure that you find out about supply and demand. Nothing else in this book brings you as great a practical reward as the contents of this chapter. Reading this chapter provides you with new insights on virtually everything concerning commerce, business and politics.

At the same time, you don't want to overdo it. The supply and demand model is a model of how markets function, but not everything in life is a market. Economics sometimes gets a bad name because it can seem that economists try to explain everything using supply and demand. As the famous English historian Thomas Carlyle once sneered, 'Teach a parrot the terms "supply and demand" and you've got an economist.'

We begin this chapter by introducing you to markets. We then explain supply and demand separately and show you how to draw and manipulate supply curves and demand curves; the demand curves capture the behaviour of buyers, whereas the supply curves capture the behaviour of suppliers. The next step is to watch the curves interact to see how markets function both when left to their own devices and when subject to government regulation or intervention.

Making Sense of Markets

In the modern economy, most economic activity takes place in *markets*, places where buyers and sellers come together to trade money for a good or service. A market doesn't have to be an actual (physical) place; in fact, many markets nowadays are fully computerised and exist only in cyberspace (online music stores, for example). But no matter what sort of institutional arrangement governs markets, they all tend to behave in the same way, which means we can study markets in general instead of having to study each one separately.

We use a very simple model called *supply and demand*, which relates the number of people producing or buying a good to its price – holding other things constant.

This model logically separates buyers from sellers and then summarises each group's behaviour with a single line on a graph. The buyers' behaviour is captured by the demand curve, whereas the sellers' behaviour is captured by the supply curve. By putting these two curves on the same graph, economists can show how buyers and sellers interact to determine how much of any particular item may be sold, as well as the price at which to sell it.

But before we get to that handy graph, we need to explain exactly where the two curves come from and how you can manipulate them to describe different sorts of human behaviour. We tackle the demand curve first, and then the supply curve.

Deconstructing Demand

People want to buy things, and economists refer to that desire as demand. When they say *demand*, economists aren't referring to pie-in-the-sky dreams or mere wishful thinking along the lines of 'I want a bajillion scoops of ice cream!' Instead, when they say *demand*, economists mean how much of something people are both *willing and able* to pay for. So although you may want a bajillion scoops of ice cream, that desire isn't a demand in the economic sense. Your *demand* is in fact three scoops because you're willing and able to buy three scoops at the price that the local ice cream shop charges.

Getting our terms straight



To be precise in terminology, we need to distinguish between two slightly different concepts. The ice-cream scoop scenario actually describes *quantity demanded*, which refers to how much you demand at a specific price *given your income and preferences*. In contrast, when an economist uses the word *demand*, he or she means the whole range of quantities that a person with a given income and preferences demands at various possible prices.

This may seem like hair splitting, but the difference is crucial and prevents us from getting confused between the two concepts: just remember that *demand* is general across all prices and preferences, and *quantity demanded* is specific to one price.

To get a better handle on the difference between these two concepts, you have to understand that economists divide everything that can possibly affect the quantity demanded into two groups: the price and everything else. The two groups have different effects.

Prices have an *inverse relationship* with the quantity demanded. In other words, the higher the price, the less people demand (if all the other things that may possibly affect the quantity demanded are held constant).



Other things that we hold constant include such important factors as tastes, preferences and incomes. For example, no matter how low the price of a ticket gets, Spurs supporters aren't going to buy a single seat to watch Arsenal, because they don't value having a seat. At the same time, however, many thousands love Arsenal so much that even if the price gets very high (and watching football is a very expensive pastime these days), they're still willing to buy a seat.

No matter how much a seat at Arsenal costs, the people who love the team have a higher quantity demanded than Spurs fans. Because this remains true for every possible price, we say that Arsenal fans have a higher demand for this particular product than Spurs fans.

Another important factor is income. As you get richer, you increase your purchases of certain goods that you've always liked and can now afford to buy in larger quantities. These goods are called *normal goods*. On the other hand, you decrease your purchases of things that you were buying only because you were too poor to get what you really wanted. These goods are called *inferior goods*. For example, new cars are normal goods, whereas really old, poorly running used cars are inferior goods. Similarly, freshly made organic salads are normal goods, whereas three-day-old discounted bread is an inferior good.

Given the complexity of variables such as preferences and income, why do economists insist on dividing everything that can possibly influence your quantity demanded into only two groups, the price and everything else? They do so for two reasons:

- ✓ When you don't separate prices from everything else, things get confusing, and economists want to concentrate on prices.
- ✓ When you translate the concept of demand into a graph and create a demand curve, prices have a very different effect than the other variables. This point is what we show you next.

Graphing the demand curve

If we add up the demands from different consumers in a market, we get the *demand curve* for the market.

A demand curve is shown in Figure 8-1. Suppose that this demand curve represents the demand for cabbages. On the vertical axis is the price of cabbages, measured in pounds sterling. The horizontal axis is the number, or quantity, of cabbages that are demanded at any given price.

You get the total revenue gained from selling a given number of units at a given price by reading off the price and quantity demanded. For example, at point *A* you sell five cabbages at a price of £2 per cabbage, and therefore you receive a total of £10.

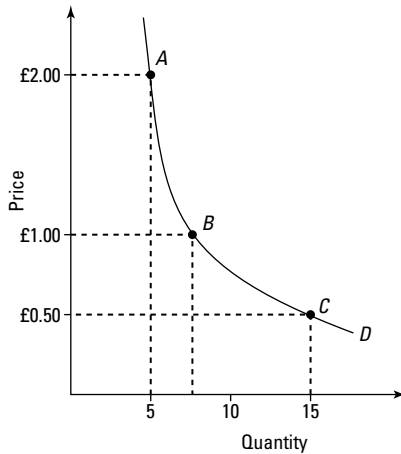
As Figure 8-1 shows, the demand curve slopes downward, reflecting an inverse relationship between the price of cabbages and the number of cabbages people want to buy. For example, consider point *A* on the demand curve. At a price of £2 per cabbage, people demand five cabbages. However, as point *B* demonstrates, if the price drops to £1 per cabbage, people demand eight cabbages. And if the price drops to only 50 pence per cabbage, they demand 15 cabbages.

Price changes: Moving along the demand curve

When you consider the relationship between the price and the quantity demanded at each price, you need to understand that increases or decreases in price simply move you along the demand curve, so that you just read off the price and quantity at a new point on the same curve.

In the previous section, we mention that economists divide all the variables that can affect demand into two groups, price and everything else. Geometrically, this division is reflected in the fact that price changes move you along the demand curve, whereas the other variables combine to determine the curve's exact location and shape.

Figure 8-1:
Demand
curves
slope
downward
because
when prices
fall, people
buy more.



For example, if people hate cabbages, you aren't going to find them buying five when the price is £2, as they do at point A in Figure 8-1. If people hate cabbages, they buy none at all, no matter what the price, and the demand curve looks very different.

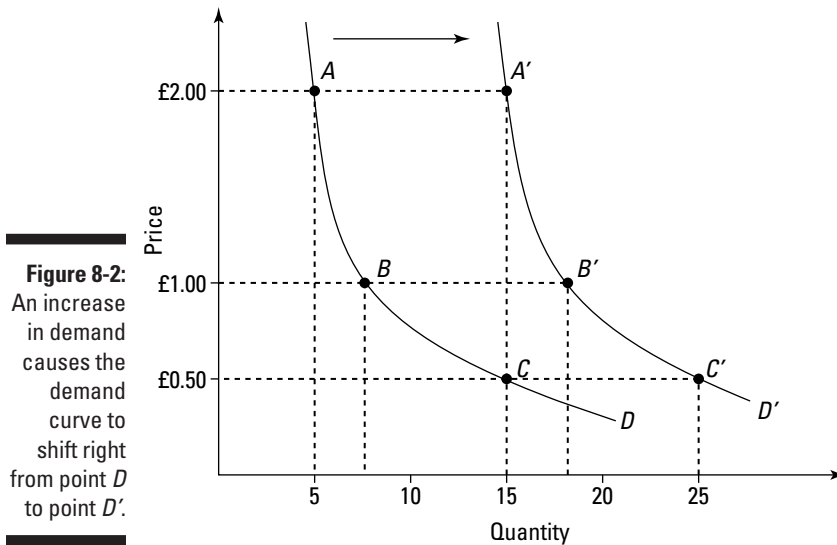
Other changes: Shifting the demand curve

Since the non-price factors determine the location and shape of the demand curve, any change in these factors causes the demand curve to shift its location.

For example, suppose that a government health study comes out saying that cabbages make people irresistible to other people. Naturally, the demand for cabbages increases, *at any given price*. Using our graphs, the effect is to shift the demand curve to the right. We illustrate this effect in Figure 8-2, where point D indicates the demand curve before the study is announced and D' indicates the demand curve after the study is announced.

Whenever a demand curve moves, economists say that a *shift in demand* has occurred. In this case, you can say that demand has increased, whereas if the curve shifted to the left, you say that demand decreased.

Making sure we're clear, in this way of describing the movements, the quantities demanded increase or decrease *while holding prices constant*. We need to emphasise this point: you *have* to distinguish between changes in quantities demanded that occur because the price changes (movements along a given curve) and changes in quantities demanded that occur because something other than the price changes (shifts of the entire curve).



To see the difference, compare point A and point A' in Figure 8-2. Both points share the same price of £2 per cabbage, but thanks to the recently released government study, people now demand fifteen cabbages at that price (point A') rather than five cabbages at that price (point A). Because the price is the same for the two points, you *know* that the change in the quantity demanded was caused by something other than price. Similarly, you can look at what happens to the quantity demanded while holding the price constant at £1: the quantity increases from eight before the study to eighteen after, moving from point B to point B' .

Remember that anything other than the price that affects the quantity demanded shifts the demand curve. In our example, a positive research study causes people to demand more cabbages. But many other factors may influence people's demand, including changes in their income or wealth and changes in their tastes or preferences. Whenever any of these non-price factors changes, the demand curve shifts left or right.

Opportunity costs: Determining the slope of the demand curve

The slopes of demand curves depend on how people view the trade-offs that changing prices force them to make. For instance, imagine that the price of a good you currently buy falls from £10 down to £9. How do you respond? Well, that depends on how you feel about the good in question relative to other goods you can spend your money on:

- ✔ You may buy a lot more of the good in question because extra units bring you a lot of happiness, and you're consequently grateful to be able to purchase them for £9 instead of £10.
- ✔ You may barely increase your buying because, although you like being able to buy the good for £9 rather than £10, extra units just don't make you all that much happier. In such a situation, the best thing about the price cut is that it frees up money to buy more of other things.

In terms of demand curves, these different reactions lead to different slopes. The person who buys a lot more when the price falls has a flat demand curve, whereas the person whose purchases barely budge when the price falls has a steep demand curve.



To make this discussion more concrete, consider Figure 8-3, where we show two separate demand curves on two separate graphs. The one on the left is your demand for sherbet lemons. The one on the right is your friend's demand for sherbet lemons.

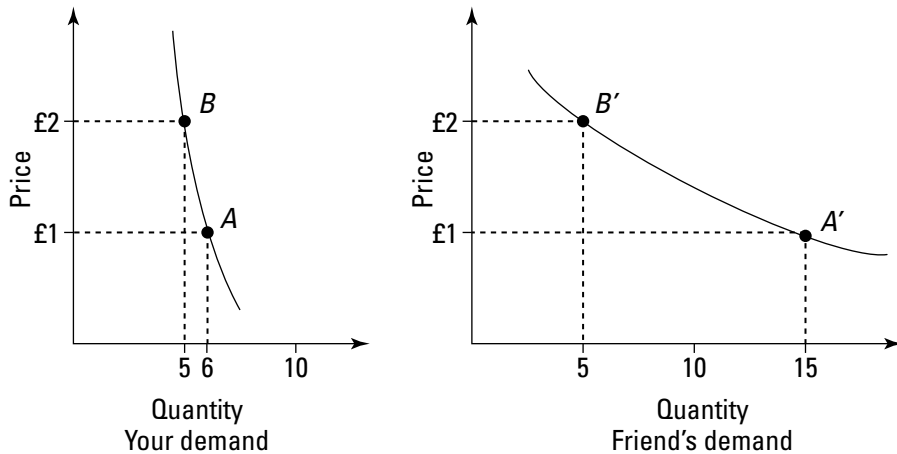


Figure 8-3:
Two
demand
curves for
sherbet
lemons.

Notice that your demand curve has a very steep slope whereas your friend's demand curve is very flat. The difference is completely the result of differences in how you react to price changes. You can see this by comparing your quantity demanded at point A with your quantity demanded at point B. Even though the price doubles from £1 per bag of delicious sherbets to £2 per bag, your quantity demanded falls only from six bags to five bags. In contrast, when the price doubles from £1 per bag to £2 per bag, your friend's quantity demanded falls hugely, from fifteen bags to only five bags.

Loosely speaking, this means that your friend is much less attached to sherbets than you are. When you see the price double, your quantity demanded barely reduces, meaning that you're willing to give up a lot of other things that you can spend the money on in order to keep buying almost as many sherbets as before.

Your friend, on the other hand, reacts very differently. Although she initially buys more sherbets than you do when the price is only £1, doubling the price causes her to cut her sherbet buying by ten bags. When the price doubles, she decides that she's better off cutting back sharply on sherbet purchases in order to spend her money on other things. In plain English, she's not nearly as attached to sherbets as you are. (But then, what can you expect from an Arsenal fan!)

Defining demand elasticity



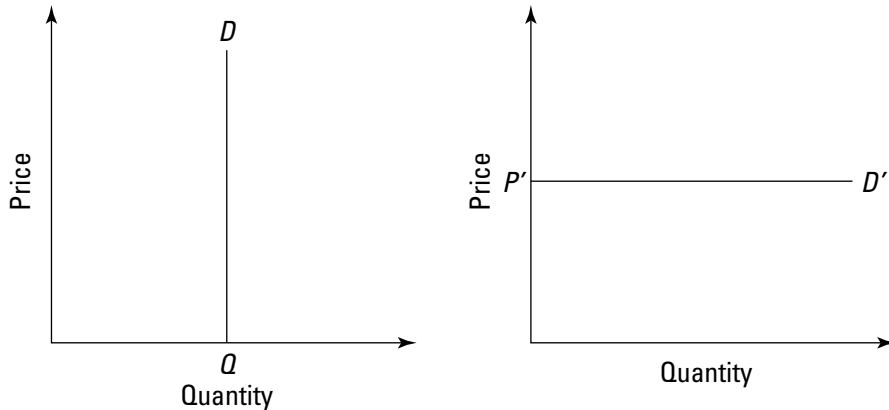
Economists have repurposed the word *elasticity* to describe how changes in one variable affect another variable. When they say demand elasticity, they mean a concept called the *price elasticity of demand* – how much the quantity demanded changes when the price changes. In Figure 8-3, your demand curve has a lot less demand elasticity than your friend's because the same change in price causes your quantity demanded to fall much less than your friend's quantity demanded.

Extreme cases of demand elasticity are illustrated in Figure 8-4 using two demand curves, the first being perfectly vertical and the second being perfectly horizontal.

The vertical demand curve, D , is said to be perfectly inelastic, because exactly Q units are demanded, no matter what the price. You may be wondering just what sort of a good has such a demand curve, and one answer is lifesaving drugs. If you need exactly Q units to keep living, you're willing to pay any price asked. Ransoms in kidnappings are also likely to be perfectly inelastic, because people are willing to pay any price to get their family members back. In fact, any good has a demand curve like this when your valuation for the good is so extreme that you're willing to pay anything for it.

On the other hand, the horizontal demand curve, D' , is said to be perfectly elastic. To understand this name, try to imagine a very gradually sloping demand curve that's almost – but not quite – horizontal. On such a very shallowly sloped demand curve, even a small change in price causes a big change in the quantity demanded. Indeed, the flatter a demand curve becomes, the greater is the change in the quantity demanded for any given price change. For instance, look at Figure 8-3 one more time. Compare how a £1 change in the price of sherbets causes a much bigger change in your friend's quantity demanded on her flatter demand curve than on your steeper demand curve.

Figure 8-4:
Two
extreme
cases of
demand
elasticity.



You can think of a perfectly horizontal demand curve as being the most extreme case of this phenomenon, so that even the tiniest change in price brings forth an infinite change in quantity demanded. That is, when prices are above P' in the right-hand graph in Figure 8-4, you buy nothing, whereas when prices are at P' or just a penny less, you buy a whole lot. (Infinite is a whole lot.)



EXAMPLE

Suppose you work for a large restaurant chain and have to buy tons of tomato ketchup. Your options are brand X and brand Y, but because they taste exactly the same, the only thing that matters is the price. Consequently, when the price of brand X is even the slightest bit lower than brand Y, you buy tons of brand X and none of brand Y. When the price of X is even slightly higher than that of brand Y, you buy tons of Y and none of X. Or to put it another way, when brand X is easily substituted with brand Y, the demand for brand X tends towards being perfectly elastic too.

In this case, an interesting thing applies when you're on the supply side of the market: if you know the elasticity, you can tell whether you're wise to raise prices. Remember that the demand curve maps a price to a quantity. Multiply that price by that quantity and you have the revenue you can make from a product being bought at that price. The result is the area of a box drawn between the demand curves and that price and quantity.

Now, if you changed the price of the good, you get another box with an area equal to the revenue to be made from selling at the new price. You can compare the size of the two boxes to tell whether you're better off at the higher or lower price. (Holding all other things constant, of course.)

Of course, perfectly elastic or perfectly inelastic demand curves aren't very common. Nearly all demand curves slope downward, meaning that moderate changes in prices cause moderate changes in quantities demanded. In Chapter 9, we explain why this is the case by looking at how consumers make trade-offs between different goods in order to maximise the happiness that they can get from spending their limited budgets. But before we get to that, we're going to introduce you to the demand curve's partner in crime, the supply curve.

Sorting Out Supply



We now move on to how economists view the supply of goods and services. The key underlying concept is that supplying things is costly. As any business owner knows, you have to pay people to supply the things you want. Even more interesting, though, is the fact that the more you want them to supply, the higher their costs of supplying each additional unit. (In Chapter 10, we explain why this situation remains true.)

Because production costs rise as you produce more output, when you want producers to make more and more, you have to pay them more and more. This fact implies that supply curves slope upward.

Graphing the supply curve



Using cabbages again as an example, imagine that a farmer named Babbage likes to grow cabbage. In Figure 8-5, we graph Mr Babbage's supply of cabbages and label that supply as *S*.

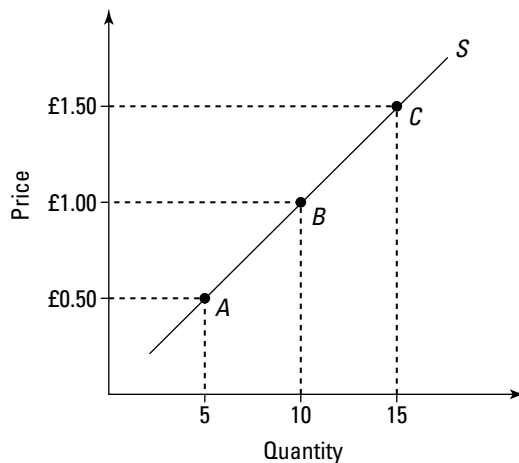


Figure 8-5:
Supply curves have an upward slope because of increasing production costs.

The horizontal axis gives the number of cabbages supplied, whereas the vertical axis gives the price per cabbage that you have to pay Mr Babbage to supply you any given number of cabbages. Thus, point *A* says that you have to pay Mr Babbage 50 pence for each cabbage if you want him to supply you with five cabbages.

Because Mr Babbage's production costs rise as he tries to grow more and more cabbages, you have to pay him £1 per cabbage if you want him to grow you ten cabbages, as shown by point *B*. And you have to pay £1.50 per cabbage if you want 15 cabbages, as shown by point *C*.



Keep in mind that the points on the supply curve don't represent the prices that Mr Babbage wants to receive for any given amount of cabbages – obviously, he wants to receive as much as he can get for each one. Instead, each amount in pounds on a supply curve represents the minimum that you can pay him and still get him to produce the desired amount. At point *A*, you can get him to produce five cabbages if you pay him 50 pence for each cabbage; if you offer him 49 pence, he refuses. Why? Because he has costs (such as land, fertiliser, labour, tractors and other things farmers need to buy), and he can cover them when he gets 50 pence for each cabbage but not 49 pence.

Separating sales price and production cost

As with demand curves, economists split all the things that can affect the quantity supplied into two groups: the price and everything else. The things that go into everything else relate to the costs of supplying the good in question.

When you see a particular supply curve, imagine that it derives from a particular production technology used by the supplier. (When an economist says *a technology*, he or she means a particular way of combining inputs to make outputs.) Because each possible technology creates its own unique relationship between output levels and costs, some technologies give rise to steeply sloped supply curves, whereas others generate fairly flat supply curves. (See Chapter 10 for the details on firms' supply curves.)

Regardless of exactly how the curve is sloped or its position, the fact that costs increase as output increases means that you need to offer a higher and higher price to the supplier when you want to obtain more units. And that is the basic reason why prices move you along supply curves. The next two sections explain these ideas in more detail.

Price changes: Moving along the supply curve

Varying the price of an item moves you along a given supply curve because the supply curve represents the minimum payment you need to give the supplier in order for him or her to supply the amount of output you want.



To see how this works, think about cabbages again. Consider what happens if you offer to pay Mr Babbage £1 per cabbage, and then you let him choose how many cabbages he wants to produce. Given his supply curve in Figure 8-5, he wants to produce exactly ten cabbages and no more, because for cabbages one to nine the cost of production is less than you're paying him. For example, consider point *A* in Figure 8-5. At point *A*, Mr Babbage's production costs are 50 pence per cabbage, meaning that if you pay him £1 per cabbage for five cabbages, he makes a nice profit. Similarly, because his cost per cabbage for producing six cabbages is also less than £1 per cabbage, he's going to want to produce a sixth cabbage. The same is true of cabbages seven, eight and nine.

At ten cabbages, Mr Babbage is indifferent, because his cost per cabbage is £1 and you're offering him £1. Mr Babbage will supply up to this point. But notice that Mr Babbage doesn't produce at point *C* if you offer him £1 per cabbage, because his cost of production is £2 per cabbage, and he loses money.

So think about the supply curve and how it responds to price changes in this way: suppliers look at whatever price is being offered and produce as many units as are profitable, but no more. Because costs rise with each additional unit produced, the only way to get suppliers to produce more is to offer them higher prices. Therefore, raising or lowering prices moves you along the supply curve as the suppliers' quantities supplied respond to changing prices.

Cost changes: Shifting the supply curve

Because a supplier's cost structure determines the location and slope of the supply curve, changes in the cost structure cause changes in the supply curve. In Figure 8-6, Mr Babbage's costs of production increase because the government imposes a new organic farming law under which he's required to grow cabbages without using pesticides. In response, he has to hire lots of extra workers to kill pests with tweezers instead of simply spraying cheap chemicals.

Because his costs of production have increased, the minimum you have to pay him to produce any given level of output also goes up. Consequently, his supply curve can be thought of as shifting upward vertically from S_0 to S_1 .

We draw the shift in Figure 8-6 to show that Mr Babbage's cost of production is 50 pence higher for each cabbage no matter how many cabbages are produced. Compare points *A* and *A'*. Before the new environmental regulation, Mr Babbage was willing to produce five cabbages if you paid him 50 pence for each cabbage. After the policy change, you have to pay him £1 per cabbage when you want him to grow you five cabbages.

Similarly, points *B* and *B'* show that before the regulation, he's willing to grow you ten cabbages if you offered him £1 per cabbage. Now, you have to offer him £1.50 per cabbage when you want him to grow ten.

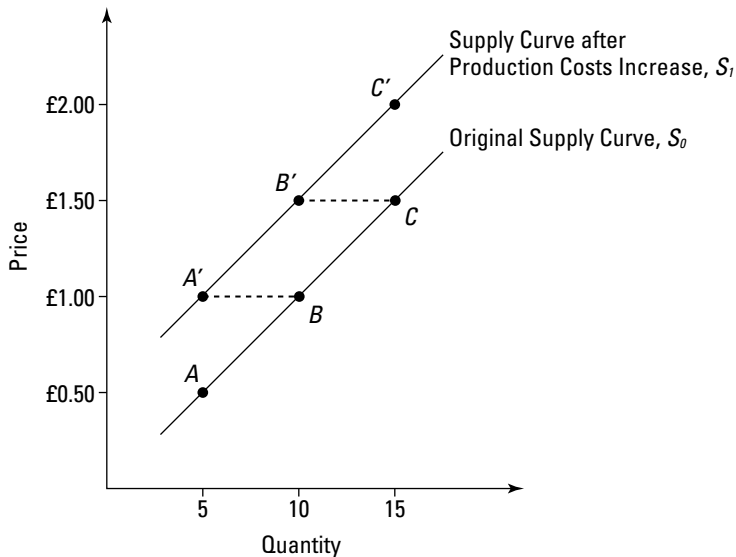


Figure 8-6:
Increased
production
costs shift
the supply
curve.

The important thing to remember is that anything that changes producers' costs structures shifts their supply curves. Things that make production more costly shift supply curves up, whereas things that lower costs shift supply curves down.

Keep in mind that thinking about supply curves as moving left and right when cost structures shift is perfectly kosher. For example, consider the quantity supplied at a price of £1 both before and after the cost increase. Before the cost increase, Mr Babbage is willing to supply you with ten cabbages for £1 each, putting you at point B on the original supply curve. But after the cost increase, he's willing to supply you only five cabbages for £1 per cabbage, putting you at point A' on the shifted supply curve. Similarly, at a price of £1.50 per cabbage, Mr Babbage was previously willing to supply you with 15 cabbages (point C), whereas after the cost increase he's willing to supply only 10 cabbages at that price (point B').

You can quite accurately say that the supply curve shifted left when costs increased. And you can quickly surmise that a decrease in costs shifts the supply curve to the right. (And you're right.)

Having two ways to interpret supply curve shifts is actually rather handy. In some situations, thinking of the shifts as right or left is easier, whereas in other cases thinking of them as up or down is easier.

Understanding extreme supply cases

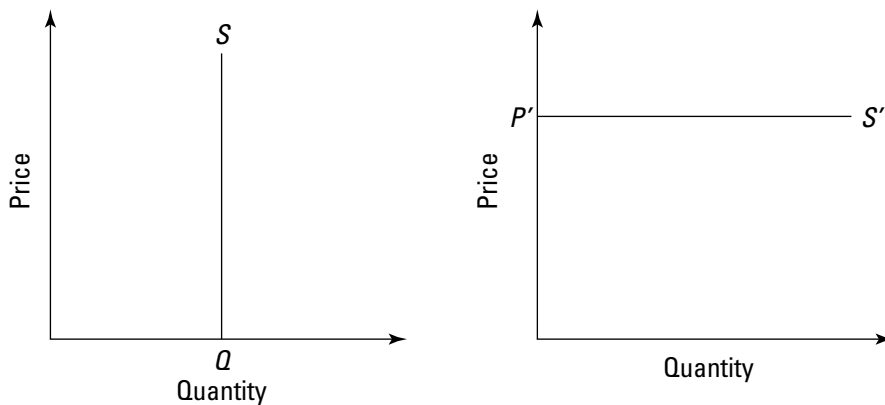
Two extreme supply curves help to illustrate how production costs and prices combine to determine the quantity to be supplied at any particular price. We illustrate these two cases in Figure 8-7.

The graph on the left shows a vertical supply curve and illustrates what economists call *perfectly inelastic supply*. The graph on the right with a horizontal supply curve illustrates what economists call *perfectly elastic supply*. We talk about each curve in the next two sections.

Paying any price: Perfectly inelastic supply

The left graph of Figure 8-7 illustrates a situation in which the price has no effect on the quantity supplied. As you can see in the graph, no matter how low or how high the price, the quantity Q is supplied. Because the quantity supplied is completely unresponsive to the price, economists call this situation *perfectly inelastic*, and supply situations that look like this are usually referred to as situations of *perfectly inelastic supply*.

Figure 8-7:
Extreme
supply
curves.



We expect you're curious about what things have perfectly inelastic supply curves. The answer is unique things that can't be reproduced. Examples include:

- ✓ **The Hope Diamond:** Because only one Hope Diamond can exist, no matter how much anyone wanted to pay, its supply curve is vertical.
- ✓ **Land:** As comedian Will Rogers said back in the early 20th century, 'Buy land. They ain't making more of it.'
- ✓ **The electromagnetic spectrum:** Only one set of radio frequencies exists, and we all have to share because making more frequencies is impossible.

An interesting thing about unique situations is that no production costs are involved. Because of this, offering the owner a price is not an incentive as it may be when you pay a producer enough to make something for you. Instead, the price serves solely to transfer the right of ownership and usage from one person to another.



A good example in this context is the mobile phone spectrum auction, which yielded the government £22.5 billion in 2000. Because the spectrum is in fixed supply and the government controls that supply, by law and convention, they were able to auction off the spectrum and received high bids. Mobile networks bought at high prices, because no other means existed to supply the spectrum.

Producing however much you want: Perfectly elastic supply

The right-hand graph in Figure 8-7 illustrates the polar opposite case, where the supply curve is perfectly horizontal. The idea here is that the supplier is producing something with non-increasing costs. No matter how many units you want, the supplier's cost remains only P' pounds to make a unit. Consequently, whether you want one unit produced or one jillion units produced, you pay only P' pounds per unit.

In the real world, perfectly elastic supply curves are rare to the point of non-existence, because production costs typically rise with output levels (as we explain in Chapter 10). However, in the virtual world they are common. For example, it costs Google a negligible amount to supply an extra search. One problem associated with that is that it proves to be very difficult to charge direct users for their searches. (We take a gander at some of the implications in Chapter 15).

Interacting Supply and Demand to Find Market Equilibrium

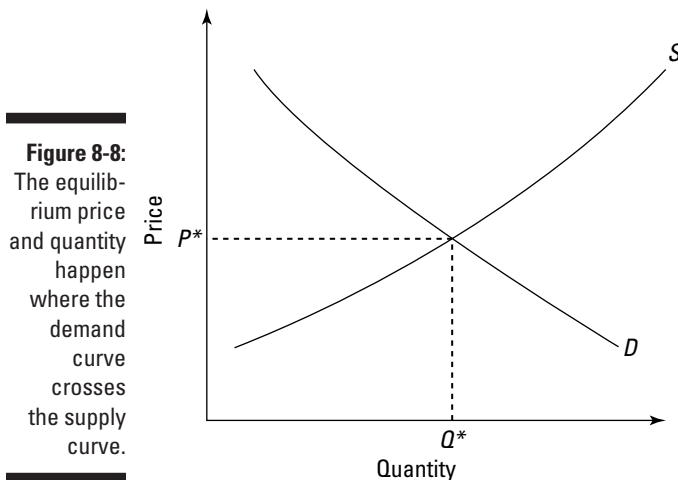
In previous sections we discuss demand and supply curves separately. Now we bring them together so that they can interact to show you how markets determine the amounts, as well as the prices, of goods and services sold.

Finding market equilibrium

In Figure 8-8, we show a demand curve and a supply curve on the same axes, labelled D and S , respectively. Remember three things about the demand and supply model when looking at this graph:

- ✓ The *equilibrium* of the supply and demand model is where the demand and supply curves cross. (Just remember: *X* marks the spot!)
- ✓ The price and the quantity where the curves cross are, respectively, how much the good or service in question costs and how much of it gets sold. This price and this quantity are known as the *market price* and the *market quantity*.
- ✓ The market price and market quantity represent a stable equilibrium such that market forces always push the price and quantity back to these values. Consequently, the market price and market quantity are also called the *equilibrium price* and the *equilibrium quantity*.

We label the market price and market quantity as P^* and Q^* , respectively. What makes this price and this quantity special is that at price P^* , the quantity that buyers demand is equal to the quantity that producers want to supply.



Put slightly differently, you can see starting at price P^* and moving to the right horizontally along the dotted line that buyers demand Q^* at that price and sellers supply Q^* at that price. Because demand equals supply, both producers and consumers are content. The consumers get exactly the quantity that they want to buy at price P^* , and the producers sell exactly the quantity that they want to sell at price P^* .

Economists call situations like these, where everybody is happy, equilibriums, because with everyone getting everything that they want, nobody is going to cause any changes.

Even more interesting is that at any price other than P^* , buyers or sellers always bring some sort of pressure to bear to return the model to the market equilibrium price and quantity. The pleasant result is that no matter where the market starts, it always ends back at equilibrium.



Before we talk more about the supply and demand model's equilibrium, you need to know one other very important thing. Notice that at the market equilibrium quantity, Q^* , the price that buyers are charged, P^* , is on the supply curve. This means that suppliers are just barely getting enough money to motivate them to supply quantity Q^* . In other words, suppliers aren't able to exploit buyers. This result tells us that capitalism isn't the same thing as pricing exploitatively as long as there is competition. Quite the contrary: if real competition exists, producers just barely earn enough money to make staying in business worth their while. (We talk much more about this subject in Chapter 11.)

Demonstrating the stability of the market equilibrium

The market equilibrium is called a stable equilibrium because no matter where the demand and supply model starts off, it always gravitates back to the market equilibrium – just so long as we don't introduce any outside forces! This inherent stability is great because it means that markets are self-correcting, and when you know where the demand and supply curves are, you know where prices and quantities are going to end up. Especially gratifying is the fact that the actions of the market participants – buyers and sellers – move the market towards equilibrium without the need for any outside intervention, such as government regulations.

We want to prove to you that the market equilibrium is indeed stable. In the next section we focus on the fact that when prices start higher than P^* , they fall down to P^* . After that, we show you that when prices start lower than P^* , they rise up to P^* . The fact that prices always move toward P^* indicates that the market equilibrium is stable.

Excess supply: Reducing prices until they reach equilibrium

In Figure 8-9, you can see what happens when you have a price like P^H that starts out higher than the market equilibrium price, P^* . At price P^H , the quantity demanded by buyers, Q^D , is less than the quantity supplied by sellers, Q^S . (We use dotted lines to show where P^H intersects the demand and supply curves.) Economists refer to such a situation as excess supply, and it can't be an equilibrium because sellers aren't able to sell everything they want to sell at price P^H .

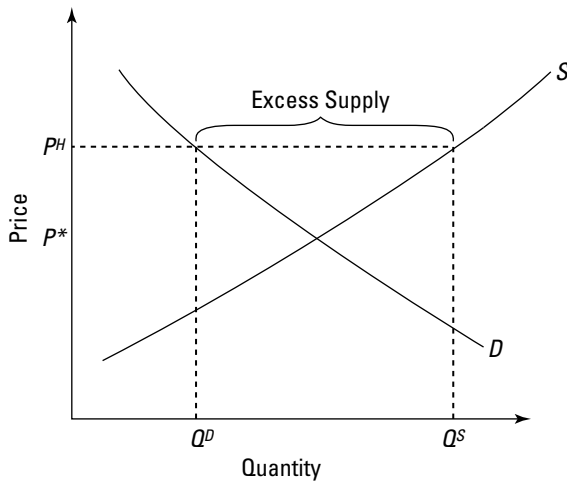


Figure 8-9:
Excess
supply.

In fact, of the total amount that sellers want to sell, Q^S , only the amount Q^D is sold, meaning that the remaining amount, $Q^S - Q^D$, remains unsold unless something is done. Well, something is done. Sellers see the huge pile of unsold goods and do what any store does when it can't sell something at current prices: they have a sale.

Sellers lower the price and keep lowering until supply no longer exceeds demand. You can see in Figure 8-9 that this means sellers keep lowering the price until it falls all the way down to P^* , because P^* is the only price at which the quantity demanded by buyers equals the quantity that sellers want to supply.

Excess demand: Raising prices until they reach equilibrium

Figure 8-10 shows a situation opposite to the one in the preceding section. The initial price, P^L , is lower than the market equilibrium price, P^* . You can see that in this case, the problem is not excess supply, but rather excess demand, because at price P^L the amount that buyers want to buy, Q^D , exceeds the amount that suppliers want to sell, Q^S .

In other words, a shortage exists that equals $Q^D - Q^S$ units. As a result, buyers start bidding the price up, competing against each other for the insufficient amount of the good.

As long as the price is less than P^* , some degree of shortage exists, and the price continues to be bid up. This means that whenever you start out with a price less than P^* , the price is pushed back up to P^* , returning the market to its equilibrium – the only place where neither a shortage nor an excess supply exists.

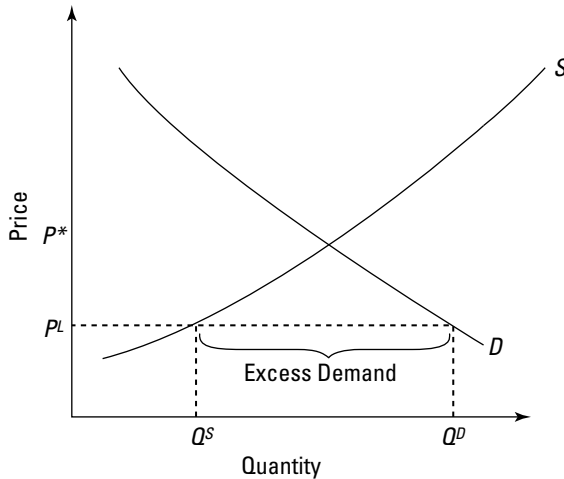


Figure 8-10:
Excess
demand.

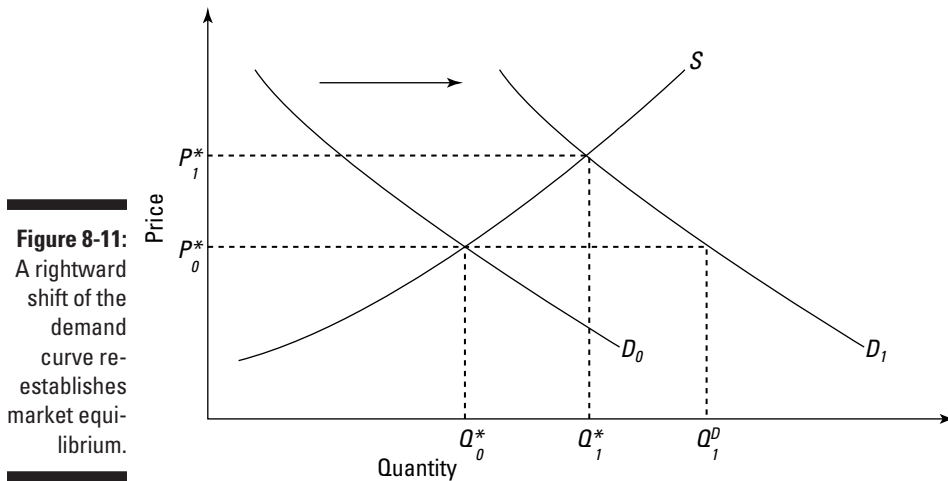
Adjusting to New Market Equilibriums When Supply or Demand Changes

As shown in the previous sections, for any given supply and demand curves, market forces adjust the market until the price and quantity correspond to where the demand and supply curves cross. When they reach that point – the market equilibrium – the price and quantity don’t change. They stay right there as long as the demand and supply curves don’t move.

In this section, we show you how prices and quantities do adjust when the demand and supply curves change. We illustrate this adjustment by showing you a demand curve shift and then a supply curve shift.

Reacting to an increase in demand

Take a close look at Figure 8-11, which shows what happens when the demand curve shifts to the right from D_0 to D_1 while the supply curve S stays the same. Before the shift, the market equilibrium price is P^*_o , and the market equilibrium quantity is Q^*_o . When the demand curve shifts to the right to D_1 , the price momentarily stays the same at P^*_o . But this price can’t last because with the new demand curve, an excess demand now exists: that is, at price P^*_o , the quantity demanded, Q^D_1 , exceeds the quantity supplied, Q^*_o .



As we discuss in the previous section, any such shortage causes buyers to bid up the price. The result is that the price rises and continues to rise until it reaches P^*_1 , the price where demand curve D_1 crosses supply curve S .

Note that when moving from the first equilibrium to the second, the equilibrium quantity increases from Q^*_0 to Q^*_1 . This result makes good sense because if demand increases and buyers are willing to pay more for something, you expect more of it to be supplied. Also, the price goes up from one equilibrium to the other because to get suppliers to supply more in a world of rising costs, you have to pay them more.

Much more subtle, however, is that the slope of the supply curve interacts with the demand curve to determine the size of the changes in price and quantity. Refer back to the perfectly vertical supply curve of the left-hand graph of Figure 8-7. For such a supply curve, any increase in demand increases the price *only*, because the quantity can't increase. On the other hand, if you're dealing with the perfectly horizontal supply curve of the right-hand graph of Figure 8-7, a rightward shift in demand increases the quantity *only*, because the price is fixed at P^* .

Thinking through these two extreme cases hammers home that in a situation like Figure 8-11, neither demand nor supply is in complete control. Their interaction jointly determines equilibrium prices and quantities and how they change if the demand curve or the supply curve shifts.

Reacting to a decrease in supply

To show you how the market equilibrium changes when the supply curve shifts, consider Figure 8-12 in which the supply curve shifts from S_0 to S_1 because of an increase in production costs. (As we discuss in the earlier section 'Cost changes: Shifting the supply curve', this increase in costs can be considered to shift the supply curve up or to the left. In Figure 8-12, we use a vertical arrow to indicate a vertical shift, but a left arrow to indicate a leftward shift is just as correct.)

The shift in supply causes the market equilibrium to adjust. The original equilibrium is at price P_0^* and quantity Q_0^* , which is the point where the demand curve D and the original supply curve S_0 cross. When production costs increase, the supply curve shifts to S_1 .

For a moment, the price remains at P_0^* . But this price can't continue because the quantity demanded at this price, Q_0^* , exceeds the quantity supplied, Q_1^S . This situation of excess demand causes the price to be bid up until reaching the new equilibrium price of P_1^* , at which price the quantity demanded equals the quantity supplied at Q_1^* .

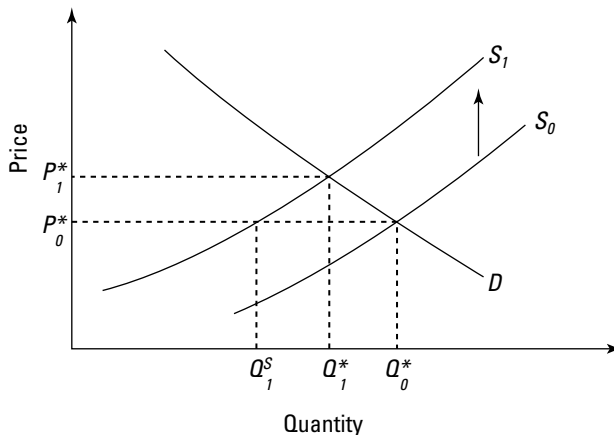


Figure 8-12:
The supply curve shifts vertically when market equilibrium resets.

When you compare this situation of increasing costs with the situation of increasing demand in the previous section, you notice that in both cases the equilibrium price rises. However, make sure to note that the equilibrium quantities go in opposite directions. An increase in demand causes an increase in equilibrium quantity, but an increase in costs causes a reduction in equilibrium quantity.



Equilibrium quantity falls because the increase in production costs doesn't just affect the producer. In order to stay in business, the producer has to pass along the cost increase. But when he or she passes the increase along, it tends to discourage buyers. The result is that the equilibrium quantity falls because some buyers aren't willing to pay the higher costs. Those who still want to buy are willing to pay the higher costs – as reflected in the increased market price.

Constructing Impediments to Market Equilibrium

Left to its own devices, a market always adjusts until the price and quantity are determined by where the demand and supply curves cross. The market equilibrium price has the great property that everyone who wants to buy at that price can do so, whereas everyone who wants to sell at that price can also do so. (The quantity demanded equals the quantity supplied.)

Remember though the market price is not always the politically expedient price, and governments often interfere in the market to prevent the market equilibrium from being reached. Such interventions happen because politically influential buyers think the market price is too high, or because politically influential sellers think the market price is too low.

Unfortunately, when the government intervenes to help the people who are complaining, it can create a whole new set of problems and, in some cases, even hurt the very people that the intervention was designed to help. To explain how this happens, we first explain price ceilings and then price floors. Price ceilings prevent prices from rising to the market equilibrium, whereas price floors keep prices from falling to the market equilibrium. (Obviously, you use only one or the other.)

Raising price ceilings

Sometimes the government intervenes in a market to ensure that the price stays below the market equilibrium price, P^* . Remember that prices below the market equilibrium normally rise, and therefore these policies are called *price ceilings*, because they prevent the price from rising as high as it may have gone if left alone. Prices hit the ceiling and then go no higher.

To see how this policy works, and the problems it creates, look at Figure 8-13, in which the price ceiling P^C lies below the market equilibrium price of P^* . To make clear that we have a ceiling above which the price can't rise, we draw a solid horizontal line starting from P^C and extending right.

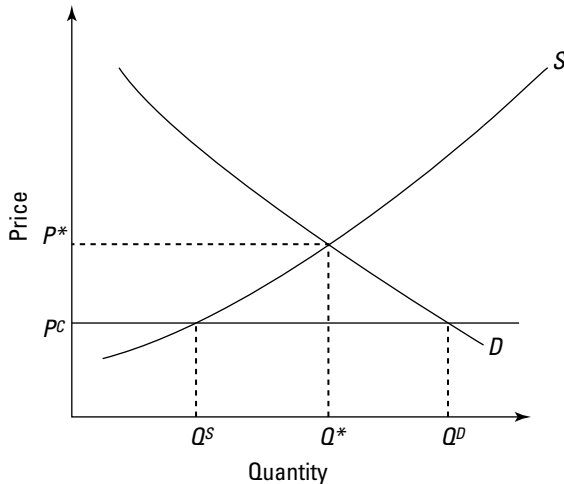


Figure 8-13:
A price ceiling.

The problem in Figure 8-13 is that at the ceiling price, the quantity demanded, Q^d , far exceeds the quantity supplied, Q^s . This situation may not look like a big problem, but you have to deal with the shortage somehow. You have to figure out a way to allocate the insufficient supply among all the people who want it. The result is that people end up queuing to get the limited supply.



The private housing market in Britain used to include price ceilings for how much a landlord was allowed to charge – a policy euphemistically referred to as rent control. This practice has largely disappeared, with the consequence that many people feel rent levels are too high (particularly students!). Although some people aren't necessarily happy about the situation, in many ways the change was an improvement. In the 1970s, when rent controls operated, people often found getting a house or flat difficult because landlords were reluctant to put their property on the market at the prevailing regulated prices. Therefore, excess demand for housing existed at that price. After the rent controls were abolished, more properties came onto the rental market, albeit at a higher price. After the impediment to supply and demand adjusting changed, the excess demand was cleared from the market to some degree. Although many people say that the housing market is still a problem in the United Kingdom, at least the reasons are different!

Propping up price floors

The opposite sort of market intervention is a *price floor*, by which the government keeps the price above its market equilibrium value. An example of this situation is shown in Figure 8-14, where the floor price, P^f , is greater than the market equilibrium price, P^* . To make it clear that prices can't fall below P^f , we draw a solid horizontal line at that price.

The problem in Figure 8-14 is that at price P^F , the quantity supplied, Q^S , is much bigger than the quantity demanded, Q^D . The normal response to such a situation of excess supply is for the price to fall. To prevent the price from falling, the government steps in and buys up the excess supply to prop the price up.

In other words, of the total amount Q^S supplied at price P^F , regular consumers demand and purchase Q^D . The remainder, $Q^S - Q^D$, must be purchased by the government. This situation doesn't sound too bad until you read about price floors in agriculture, which proponents usually refer to euphemistically as *price supports* (as in, 'You poor thing! All you need is a little support!').

Price supports generate huge piles of crops that nobody wants to buy. For example, the European Union's common agricultural policy in effect acts as a price floor. Thanks to the policy of supporting European farmers, Europe produces mountains of unwanted butter, and lakes, if not oceans, of undrinkable wine.

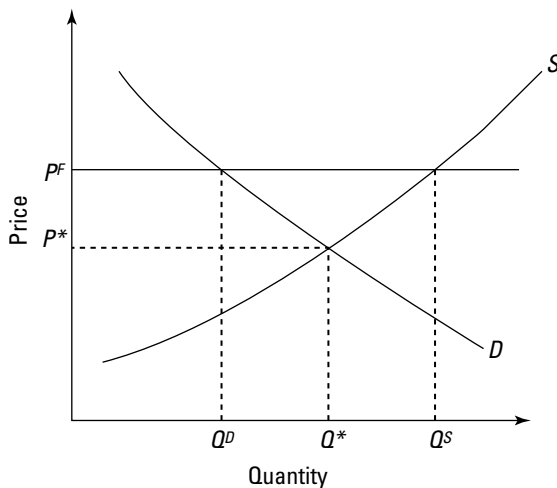


Figure 8-14:
A price floor.

Therefore, as regards both price ceilings and price floors, remember that great mischief is typically caused when you interfere with the markets. This reasoning doesn't mean that governments should *never* intervene in the operation of a market, but rather that they need to be smart enough not to intervene in ways that they know are going to lead to perverse results. (Of course, from our own individual, selfish and short-sighted perspective, we actually hope that the government decides to support the salaries of academic economists. We'd love to end up getting paid not to teach.)



First World supports, Third World suffers

One perverse result of agricultural price supports in rich countries like the United States and the nations of the European Union is the great damage those supports inflict on developing nations. For example, the US price of sugar is three times the world price because the United States restricts imports of cheaper foreign sugar in order to help US food production giants. The result is that thousands of poor Third World farmers are left without a livelihood, instead of making a living selling sugar to Americans and Europeans.

Even worse is what the United States and Europe do with some of the many tons of excess agricultural products that pile up due to the agricultural price supports. Not wanting to sell the excess in the US or EU, and thereby depress

US or EU prices, our governments often send the stuff free to developing countries as food aid. That sounds nice and friendly, but when all that free wheat hits Nigeria, Nigerian farmers are put out of business. You can see what the consequences may be for the local economy.

Interfering with markets can therefore be a very bad thing: unexpected side effects usually end up hurting people that the policy isn't expected to harm. Furthermore, such policies are also typically inefficient, costing the losers more than they benefit the winners. When you see a policy like this, it's worth thinking through very carefully where the hidden costs are falling. Although we like to apply out of sight, out of mind to them, the hidden costs can be very real and add up significantly.

Chapter 9

Getting to Know *Homo Economicus*, the Utility-Maximising Consumer

In This Chapter

- ▶ Studying how people maximise their happiness
 - ▶ Taking account of diminishing marginal utility
 - ▶ Watching how people weigh alternatives
 - ▶ Choosing exactly the right amounts within a limited budget
-

This chapter gets behind the demand curve (which we introduce in Chapter 8) by showing you how people come to choose the things they choose. This decision-making process is very important because human wants are what drive the economy. Firms don't randomly produce goods and services; they produce the things that people want to buy and are spending money on.

The thing that makes studying this process hard is the fact that people have so many different things on which they can spend their money. If an economist were asked to research how you're going to spend £100 in a shop that sells only blueberry muffins, his job wouldn't be very hard. What's impressive is that economists have come up with a way to explain how you're going to spend £100 in a shop that has hundreds or even thousands of items for sale.

Even more impressive is the fact that an economist can explain not only which items you're going to buy, but also how many of each you're going to buy. In other words, economists can explain not just *what* you demand, but also the *quantities* you demand, which is where demand curves come from.

We start the chapter by discussing *utility*, which is how economists measure human happiness. Economists assume that people act in ways that maximise their happiness, but our actions are constrained, especially by limited budgets. We explain how people navigate these constraints to get the most happiness possible given the limits involved. Finally, we show how these decisions underlie and explain the slope and position of demand curves.

Knowing the Name of the Game: Constrained Optimisation

Later in this chapter we discuss *how* people choose what to buy when they buy. But first we focus on *why* they must choose.

People need to make choices because their means for satisfying their wants are limited. You never have enough money or time to do everything you desire. Consequently, you need to choose wisely to get the most happiness out of the limited resources that you do have.

Economists and engineers refer to problems of this sort as *constrained optimisation problems*, because people are trying to optimise their happiness given the fact that they're constrained by their limited resources. The rest of this chapter shows you how economists model the way that people go about solving their everyday constrained optimisation problem: working out how people decide to best spend their limited incomes on available goods and services – choosing not only which things to buy, but also how much of each.

Finding a Common Denominator to Measure Happiness: Utility

In order for people to choose between the exceedingly different goods and services available in the economy, they must have a way of comparing them all. Comparing costs is pretty easy; you just compare prices. But how do you compare the benefits of various goods and services? How do you assess whether spending £20 on Swiss chocolate bars or on a cotton shirt is better? In what ways are chocolate and shirts even comparable?



Obviously, people do manage to make the comparison and rank the two options. Economists imagine that people do this by assigning a common measure of happiness to each possible thing they can buy and use. Economists call this common measure of happiness *utility*, and they imagine that if they

were able to get inside your brain and measure utility, they'd do so using a unit that they very uncreatively call a *util*. However, making such strong claims about how people measure happiness is not necessary.

Some people very naturally object to assigning specific numbers of utils to different things – for example, 25 utils to the pleasure associated with eating a brownie, or 75 utils to the pleasure associated with watching a sunset. (Although the philosopher Jeremy Bentham did come up with a scheme for doing just that.) Making such specific assignments is called *cardinal utility* (like cardinal numbers: 1, 2, 3 . . .). The objections to cardinal utility centre on doubts about whether people even make such assessments – after all, how many utils do you think you receive from a sunny day or a baby's smile?

A much less objectionable thing to do is to think in terms of *ordinal utility*, a system in which you simply rank things. For example, instead of saying that the sunset has a utility of 75, which makes it preferred to the brownie with a utility of 25, you can simply say that sunsets are preferred to brownies. This system has a much more intuitive feeling for most people and eliminates the need to try to measure things using the imaginary unit called the util.

Even better, it's been proven mathematically that you can describe the same human choice behaviour using ordinal utility that you can using cardinal utility, which means that economists don't *have* to use cardinal utility.

But we're going to anyway!

Why? Because using the cardinal utility system is a much easier way to explain the crucial concept of *diminishing marginal utility*. You can also explain diminishing marginal utility using the ordinal system, but the maths is a bit scary, and we usually leave this method for real geeks. So although it may seem a bit unrealistic, the cardinal utility system really is the best way to convey this incredibly important idea.

Getting Less from More: Diminishing Marginal Utility

People get bored even with things they like and get tired of repetition and sameness. Economists have to take account of this tendency when studying how people choose to spend their money.



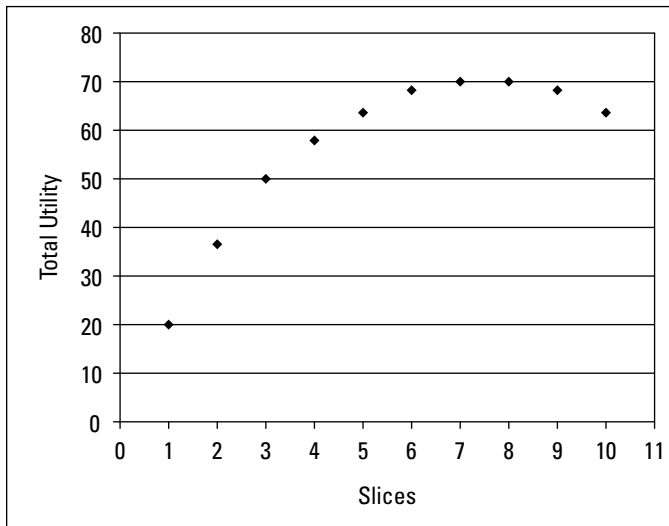
If you love your pizza, and you haven't had any pizza in a long time, you get a great big rush of utility from eating a slice. The melted cheese, the basil and garlic in the sauce, and the familiar warmth in your mouth all make you very, very happy. But eating that first slice dampens the thrill of pizza, and so if you eat a second slice, it may taste good, but not as good as the first. And if you

have a third slice, that's not as good as the second. And by the time you get to your third 13-inch stuffed-crust slice of pizza, you're likely to start to get sick and experience pain rather than pleasure.

This phenomenon isn't limited to pizza; it applies to nearly everything. Unless you're addicted to something, you get tired of it the more you have it, and each additional unit brings you less happiness than the previous unit.

To make this phenomenon clearer, look at Figure 9-1, which shows the cumulative total utility of a pizza lover who eats more and more slices of pizza. For example, total utility after eating one slice of pizza is 20 utils, after eating two slices, 36 utils and after three slices, the total utility is 50 utils.

Figure 9-1:
The total utility of a pizza connoisseur who eats more and more slices of pizza.



When you look at these numbers, you notice that the extra utility each additional slice brings is decreasing:

- ✓ **First slice:** Total utility increases by 20 utils, from 0 to 20 utils.
- ✓ **Second slice:** The increase is only 16 utils; total utility increases from 20 utils to 36 utils.
- ✓ **Third slice:** Total utility increases only 14 utils, from 36 to 50.



Economists refer to this phenomenon as *diminishing marginal utility* because the extra utility, or *marginal utility*, that each successive slice brings decreases relative to the marginal utility brought by the previous slice. (Remember that we're interested in only the extra utility from each extra slice, the marginal

utility.) Diminishing marginal utility is simply a reflection of the fact that people get fed up or bored with things. Or, in the case of food and drink, their appetite decreases with each unit they consume.

Look at what happens in Figure 9-1 after slice number eight. Total utility actually goes down, because slice number nine can make even the most rabid pizza lover feel a little sick. Add on slice number ten, and total utility falls again.

This decrease in total utility implies that marginal utility must be negative for slices nine and ten. Look at Table 9-1, which gives both the total and marginal utilities for each slice. As you can see, the data matches Figure 9-1 and shows that although total utility increases for slices one through seven, it stalls at slice number eight and falls for slices nine and ten.

Table 9-1 Total and Marginal Utility of Eating Ten Slices of Pizza

<i>Slice</i>	<i>Total Utility</i>	<i>Marginal Utility</i>
1	20	20
2	36	16
3	50	14
4	58	8
5	64	6
6	68	4
7	70	2
8	70	0
9	68	-2
10	64	-4

The right column shows the diminishing marginal utility that comes with eating more and more slices of pizza, because the marginal utility that comes with each additional slice is always less than that of the previous slice. Specifically, although marginal utility is 20 utils for the first slice, it falls to 0 utils for slice eight and then actually becomes negative for slices nine and ten because eating them makes just about anyone ill.

In Figure 9-2, we plot out the marginal utility for each slice of pizza. You can see quite clearly from the downward slope of the points that marginal utility diminishes as one eats more and more slices of pizza.

You have to be careful not to confuse diminishing marginal utility with *negative* marginal utility. As you see in Table 9-1 and Figure 9-2, diminishing marginal utility for all slices of pizza starts with the second, because each successive slice has a smaller marginal utility than the previous one. But the marginal utilities are still positive for all slices up to slice seven, and they become negative only for slices nine and ten.

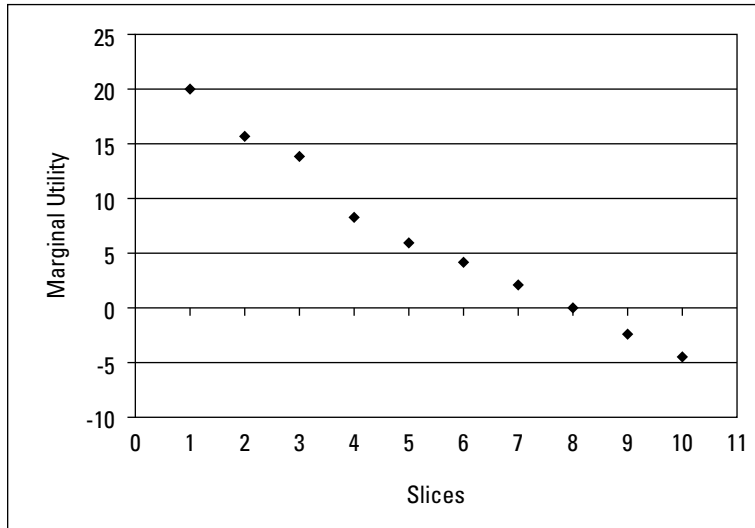


Figure 9-2:
The marginal utility derived from each slice of pizza.

That fact implies that you enjoy eating every slice up to and including the seventh slice because doing so brings you an increase in utility (happiness). So don't think that just because marginal utility is diminishing for a particular slice, you don't want to eat it. Marginal utility can be diminishing but still positive. The only slices you want to avoid outright are the ninth and tenth.

Choosing among Many Options When Facing a Limited Budget

The phenomenon of diminishing marginal utility makes studying human choices very interesting because whether you prefer chocolate ice cream or vanilla ice cream can't be determined in the abstract. Instead, your preference depends on what you've already eaten.



If you haven't had any ice cream for months and you're asked whether you want chocolate or vanilla, you may choose chocolate. But if you're asked whether you want chocolate or vanilla after you've just eaten a gallon of chocolate, you're probably going to say vanilla because you've already more than satisfied your chocolate cravings.

So the answer to the question 'Chocolate or vanilla?' isn't as straightforward as it seems. Your preferences exhibit diminishing marginal utility, and even something that you normally like a lot doesn't bring you much marginal utility (additional happiness) if you've just indulged in it a lot.

This fact ends up leading to a very simple formula about how people make decisions when faced with limited budgets. But before we state the formula, we give you an example that helps explain it.

Trying to buy as much (marginal) utility as you can



In this example, you have £10 to spend and, because you're going to the local student bar, the only two things you can spend the money on are pints of beer and slices of pizza. You're now thinking about how to best spend your £10, and the intelligent thing to do is to think in terms of buying up as much utility as you can with your limited budget. Both beer and pizza make you happy, but your goal isn't just to be happy; you want to be as happy as possible given your limited budget. So you want to make sure that every pound buys you the maximum possible amount of utility.

Keep in mind that you don't care where utility comes from. One util from beer makes you just as happy as one util from pizza; all you care about is buying up as many utils as possible.

To do that, the key concept turns out to be the price of utility. Beer and pizza clearly have prices measured in pounds, but what is the price of a util?

Well, it depends. Take a look at Table 9-2. The first three columns repeat the data from Table 9-1 that give total and marginal utilities for ten slices of pizza. But the final two columns include new data and are labelled 'MU per pound at £1 per slice' and 'MU per pound at £2 per slice' respectively. (*MU* stands for *marginal utility*.)

<i>Slice</i>	<i>Total Utility</i>	<i>Marginal Utility</i>	<i>MU per Pound at £1 per Slice</i>	<i>MU per Pound at £2 per Slice</i>
1	20	20	20	10
2	36	16	16	8
3	50	14	14	7
4	58	8	8	4
5	64	6	6	3
6	68	4	4	2
7	70	2	2	1
8	70	0	0	0
9	68	-2	-2	-1
10	64	-4	-4	-2

What we've done in these last two columns is to calculate how much it costs to get some additional happiness (marginal utility) if buying slices of pizza is the way you're getting it.

Consider the fourth column, which assumes that each slice of pizza costs £1. If you buy one slice, it brings you a marginal utility of 20 utils at a cost of £1. So the MU per pound of the first slice is 20.

But now consider spending a second pound to buy a second slice of pizza. Because that second slice brings with it a marginal utility of only 16 utils, the MU per pound spent here is only 16. And because diminishing marginal utility continues to decrease the marginal utility of each additional slice of pizza, each additional pound you spend buys you less additional utility than the previous pound.

The final column of Table 9-2 shows you that the MU per pound that you get from pizza depends on how much each slice of pizza costs. If pizza costs £2 per slice, each pound spent brings you less marginal utility than when pizza cost only £1 per slice.

For example, because each slice now costs £2, when you buy the first slice and it brings you 20 utils, you're getting only 10 utils per pound spent. Similarly, although the second slice still brings you 16 additional utils of happiness, because it now costs you £2 to get those utils, your MU per pound is only 8 utils.

In Table 9-3, we give you the same sort of information as in Table 9-2, but this time for total utility, marginal utility and MU per pound when drinking beer that costs £2 per pint.

<i>Pint</i>	<i>Total Utility</i>	<i>Marginal Utility</i>	<i>MU per Pound at £2 per Pint</i>
1	20	20	10
2	38	18	9
3	54	16	8
4	68	14	7
5	80	12	6
6	90	10	5
7	98	8	4
8	104	6	3
9	108	4	2
10	110	2	1

As you can see from the third column, you exhibit diminishing marginal utility with regard to beer, as your MU for each beer falls from 20 utils for the first pint down to only 2 utils for the tenth pint. (Wow, someone can really put it away!) As a result, your MU per pound spent in the fourth column falls from 10 per pound for the first pint down to only 1 per pound for the last pint.

Allocating money between two goods to maximise total utility

Tables 9-2 and 9-3 show you how much utility you can get by spending money on pizza or beer. The trick now is to see how you can get the most possible utility for your limited budget of £10.

As a first attempt, consider the two extreme options: blowing all the money on pizza or blowing all the money on beer. (Pizza costs £1 per slice, and beer costs £2 per pint.)

If you spend all £10 on pizza, you can buy ten slices of pizza, which gives you a total utility of 64 utils. On the other hand, if you spend all £10 on beer, you can buy five pints at £2 each and thereby get 80 total utils. If these were your only two options, you'd clearly prefer to spend all your money on beer because it brings you more utils than buying only pizza.

However, you can do a much better thing. You can get even more total utility if you wisely mix up your consumption a bit and spend some of your money on beer and some on pizza.

If you get lost the first time you go through the next bit, don't panic! The key thing to remember is that you're not interested in the item or thing you buy, but in the happiness or utility you get from it.



The way you get the most utility possible out of your £10 is simple: take each of the ten pounds in turn and spend it on whichever good brings more utility. Don't think of your task as buying slices of pizza or pints of beer, but bear in mind that your job is buying utility. For every pound spent, you want to buy as much utility as possible, and you don't care whether that utility comes from beer or pizza.

The only thing complicating this process of spending each pound on whichever good brings the most utility is the fact that you have diminishing marginal utility for both beer and pizza, meaning that the amount of utility you're able to buy with each extra pound spent depends on how much beer or pizza you've already bought. But given the information in Tables 9-2 and 9-3, you can work out the best thing you can do with each pound:

- ✔ **Pound 1:** How do you spend your first pound? From the fourth column of Table 9-2, you can see that if you spend that pound on pizza, you can buy 20 utils of utility. On the other hand, the fourth column of Table 9-3 tells you that if you spend that first pound on beer (along with a second pound because pints cost £2), you get only 10 utils of utility. So, the obvious thing to do with the first pound is to buy pizza rather than beer.
- ✔ **Pound 2:** If you use your second pound to buy a second slice of pizza, you get 16 utils of utility. If you buy beer with that second pound (along with a third pound because the price of a pint is £2), you get only 10 utils for that second pound because it's going to be spent on buying the first pint. So once again, you're better spending this pound on pizza rather than beer.
- ✔ **Pound 3:** You also want to spend the third pound on pizza rather than beer because you get 14 utils of marginal utility rather than 10 utils. (Remember, this pound would buy the *first* pint of beer, which brings 10 utils of utility.)

- ✔ **Pounds 4 and 5:** At pound number four, everything changes, because if you spend a fourth pound on pizza, it brings 8 utils. However, if you spend that fourth pound (along with the fifth pound) on a pint, you get an MU per pound of 10 utils (for each of those pounds). So, you're best spending pounds four and five on buying the first pint of beer.
- ✔ **Pounds 6 and 7:** You should also spend pounds six and seven on beer, because you get an MU per pound of 9 utils for your second pint, whereas you get only 8 utils if you spend the sixth pound on a fourth slice of pizza.
- ✔ **Pounds 8, 9 and 10:** For pound number eight, the MUs per pound are tied. If you use this pound to buy a fourth slice of pizza, you get 8 utils. You get the same by spending the pound on a third pint of beer. So what you should do is spend your last three pounds on buying a fourth slice of pizza and a third pint of beer.

In Table 9-4, we list how to spend each of your ten pounds. Notice that the total utility you can purchase with your ten pounds is 112 utils. That's much better than the 64 utils you get spending all the money on pizza or the 80 utils you get spending it all on beer. By spending each pound in sequence on whichever good brings the most utility, you've done much better than spending the money on only one good or the other.

Table 9-4 How to Spend Each Pound Optimally on Pizza and Beer

<i>Pound</i>	<i>Good Chosen</i>	<i>MU per Pound</i>
1	Pizza	20
2	Pizza	16
3	Pizza	14
4	Beer	10
5	Beer	10
6	Beer	9
7	Beer	9
8	Pizza	8
9	Beer	8
10	Beer	8
Total utils		112

Also notice that you end up buying four slices of pizza and three pints of beer. Given this budget and these prices, your quantity demanded of pizza is four slices and your quantity demanded of beer is three pints. The process of maximising utility is also the basis of demand curves and the relationship between quantity demanded and price. (We discuss demand curves in Chapter 8 and return to them later in this chapter, in the section ‘Deriving Demand Curves from Diminishing Marginal Utility’.) In the next section, we present the magic formula for choosing where to spend your money in any situation.

Equalising the marginal utility per pound of all goods and services

In the previous section, we go through a rather tedious process to determine how to best spend £10 on beer and pizza. Making these decisions doesn’t always take so long. In this section, we explain a simple formula that guides people to maximise the total utility they can get out of spending any budget, no matter how many goods are available to choose from or how much they each cost.

To keep things simple, we begin by showing you the version of the formula that applies to deciding how to best spend your budget when you have only two goods or services to choose from. When you get the hang of the two-good version, the multi-good version is effortless.

We call the two goods X and Y . Their respective prices are P_X pounds for each unit of X and P_Y pounds for each unit of Y . Also, their respective marginal utilities are MU_X and MU_Y . The formula looks like this:

$$\frac{MU_X}{P_X} = \frac{MU_Y}{P_Y} \quad (1)$$

This equation means that if a person has allocated their limited budget optimally between the two goods, the marginal utilities per pound of X and Y are equal at the optimal quantities of X and Y .

This relationship holds true in the example in the previous section. Look back at Table 9-4. When you optimally spend your £10 on beer and pizza, the optimal amounts of each are four slices of pizza and three pints of beer. From the third column of Table 9-4, you can see that marginal utilities per pound for the fourth slice of pizza and the third beer are indeed equal at 8 utils per pound, just as the formula in equation (1) dictates.

Seeing why the marginal utilities per pound must be equal

In this section, we demonstrate *why* marginal utilities per pound have to be equal if you want to maximise your utility when spending a limited budget. If marginal utilities per pound aren't equal, you want to keep rearranging your purchases until they are. The examples in this section show you why.

First, imagine that you choose some other quantities of each good, so that for the final unit of X and the final unit of Y that you purchase:

$$\frac{MU_x}{P_x} > \frac{MU_y}{P_y} \quad (2)$$



For example, let pizza be X and beer be Y . From Tables 9-2 and 9-3, you can see that if you purchase four pints of beer and two slices of pizza, the MU per pound for the fourth pint of beer is 7 utils, while the MU per pound for the second slice of pizza is 16 utils. Clearly, the MU per pound of pizza is much bigger than the MU per pound of beer if you spend your limited budget in this way.

But this way of spending your budget isn't optimal. The reason is that the money you're spending on what is currently the final unit of X (pizza) buys more marginal utility than the money you're currently spending on the final unit of Y (beer). If you can get more utility by spending a pound on X than you can on Y , take money away from spending on Y in order to spend it on X . And as long as the inequality in equation (2) holds true, continue to take money away from Y in order to increase spending on X .

Consider a more extreme example. Suppose that you spend all £10 buying five pints of beer. You can see from Table 9-3 that the marginal utility per pound of the last pound spent on beer is only 6 utils. By contrast, if you took that pound away from beer and used it to buy a first slice of pizza, the pizza brings you 20 utils (see Table 9-2). Clearly, you need to reduce your beer buying in order to increase your pizza buying.

Continue to buy fewer beers and more pizza until you arrive at the combination of four slices of pizza and three beers. That is, rearrange your spending until the marginal utilities per pound of both beer and pizza are equal, as in equation (1).

The same rule applies if you start out spending all your money on pizza. If you buy ten slices of pizza, you can see from Table 9-2 that the marginal utility of the tenth slice is actually -4 utils. Meanwhile, the marginal utility per pound of the first pound spent on beer is 10 utils. Clearly you need to take money away from pizza and use it to increase your purchases of beer.

Inflation and allocation in the real world

An interesting thing to notice when you stare at equation (1) or equation (3) in this chapter is that if all the prices in the denominators were suddenly to go up by the same multiple, all the equalities remain intact, meaning that people still choose to buy the same amounts of every good. In other words, if inflation suddenly exactly doubled all prices, people still choose to buy exactly the same quantities of everything as they did before.

People feel that logically, if your income doubles at the same time that the prices of everything you buy double, nothing has really changed. You can still purchase exactly the same quantities of goods and services as you used to purchase before the inflation. And because those quantities were the ones that were maximising your utility before, they still maximise your utility now. As a result, you may mistakenly conclude that inflation doesn't matter.

But in Chapter 5, we tell you about the great horrors of inflation. These horrors are caused by the fact that you never, in real life, see *perfect* inflation like the one we just described, in which the prices of all goods and services go up by exactly the same amount and at exactly the same time!

Instead, what happens is that the prices of different goods and services go up at different rates, so the fractions in equations (1) and (3) are thrown completely out of whack because their denominators change at different rates. When that happens, people start drastically changing their quantities demanded in an attempt to re-establish equality between all their marginal utilities per pound. As they do this, chaos results; some firms find demand suddenly falling for their products, whereas others find it suddenly rising.

So don't let equations (1) or (3) make you think that inflation doesn't matter in the real world. It does.

Applying the formula to multiple goods and services

You want to remember the rule represented in equation (1). It simply says that in order to maximise total utility, you need to rearrange your purchases so that for the final units of each good, the marginal utilities per pound are equal. If that isn't true, one of the goods offers you a higher amount of happiness for each pound spent, meaning that you want to rearrange your purchases to spend more on that good. Only when equation (1) holds do you not want to rearrange any more, because neither good offers you more happiness per pound than the other.

Equation (1) can also be generalised to apply to many goods. For example, in the case of three goods, you arrange your buying so that for the last unit of each of the three goods X , Y and Z :

$$\frac{MU_x}{P_x} = \frac{MU_y}{P_y} = \frac{MU_z}{P_z} \quad (3)$$

If any of the three goods has a higher marginal utility per pound than the others, you rearrange your purchases to buy less of the others and more of that good. And you keep rearranging until equation (3) holds true.

To keep things dead simple (because life's complicated enough without economists making it more so), we usually treat the example as being applicable to two goods, X and Y . If we want to deal with a multiplicity of goods that we may be interested in buying, we treat them as good X and *everything else*.

Deriving Demand Curves from Diminishing Marginal Utility



Diminishing marginal utility is one reason that demand curves slope downward. You can get a hint of this from Figure 9-2, where you see that the marginal utility that comes with each successive piece of pizza decreases. If your goal is to use your money to buy up as much utility as possible in order to make yourself as happy as possible, you're willing to pay less and less for each successive piece of pizza, as each successive piece of pizza brings with it less utility than the previous piece.

However, Figure 9-2 is not a demand curve, for two reasons:

- ✓ It doesn't take into account the effect that prices have on the quantity demanded.
- ✓ It looks at only one good in isolation, whereas the quantity demanded of a good is determined by finding the solution to the more general problem of allocating a limited budget across all available goods in order to maximise total utility.

In other words, you can't look at each good in isolation. How much of it you want to buy depends not only on its price, but also on the prices of everything else and how their marginal utilities vary as you buy more or less of them.

Seeing how price changes affect quantities demanded



In the example we use in this chapter, you need to decide how to best spend £10 when your choices are slices of pizza or pints of beer. Now we make one change to that example: imagine that pizza now costs £2 per slice rather than £1 per slice. We want to show you how this price change affects the quantity demanded of both pizza and beer.

The changes in quantities demanded result from the fact that the new, higher price of pizza reduces the marginal utility per pound for pizza. Doubling the price of pizza means that the marginal utility per pound spent on each slice of pizza is exactly half of what it was before. You can see this by comparing the fourth and fifth columns of Table 9-2. Because the increase in price lowers the marginal utility that each pound spent on pizza buys, naturally this affects where you spend your limited budget of £10.

As you may expect, a higher price of pizza leads you to eat less pizza and drink more beer. You can prove this to yourself by spending, in order, each of your pounds so that you buy whichever good has the higher marginal utility. (The section 'Allocating money between two goods to maximise total utility' earlier in the chapter walks you through the process.) The results of doing so are summarised in Table 9-5.

Table 9-5 Optimally Spending Your Budget When Pizza Costs £2

<i>Pound</i>	<i>Good Chosen</i>	<i>MU per Pound</i>
1	Pizza	10
2	Pizza	10
3	Beer	10
4	Beer	10
5	Beer	9
6	Beer	8
7	Beer	8
8	Beer	8
9	Pizza	8
10	Pizza	8
Total utils		90

By comparing Table 9-5 with Table 9-4, you can see that raising the price of pizza from £1 to £2 has reduced the quantity of pizza you'd order, though since you can't divide slices or pints in this model it doesn't yet affect the quantity demanded of beer. You can now only afford two slices of pizza and three pints of beer. If we were to raise the price of pizza further, you'd reallocate so that you'd only buy one slice of pizza and four pints of beer. At £3 per slice, you'd be completely out of the pizza market altogether, as the utility gained from spending all £10 on five pints of beer would be higher than the utility gained from consuming any pizza. In other words, you'd substitute towards the cheaper good to get yourself as much utility as possible overall. Economists call this the *substitution effect* of a price change.

To see why, think about it like this. We've given you a choice of two things to spend your money on, and because you've decided to choose your purchases so as to get as much utility as possible, you'll substitute towards the good that gives you the best bang for your buck in terms of maximising your utility. So as the price of one goes up and you get less utility per pound from buying it, you'll substitute *towards* the good that gives you more utility per pound spent.



The increase in the price of pizza has also made you poorer in the only sense that really matters: you're less happy. Due to the price increase, the total number of utils that you can purchase with your £10 budget has fallen from 112 down to only 90. Despite rearranging your quantities consumed of beer and pizza to make the most of the new situation, the price increase still hurts you overall. This is what's called the *income effect*.

Graphing the price and quantity changes to form a demand curve

You can use the information about how your quantity demanded changes when price goes up to plot out two points on your demand curve for pizza: four slices demanded at a price of £1, and two slices demanded at a price of £2. In Figure 9-3, we plot these two points and sketch in the rest of the demand curve. As you look at the figure, keep in mind two things:

- ✓ The downward slope of the pizza demand curve derives in part from the diminishing marginal utility of pizza, but . . .
- ✓ As the price of pizza changes, the quantity demanded of pizza doesn't change in isolation; it changes as the result of rearranging the quantity demanded of both beer and pizza in order to maximise total utility.



Demand curves for individual goods aren't made in isolation. Certainly, a relationship exists between a good's price and its quantity demanded. However, when the good's price changes, that change affects the entire budgeting decision – not just for that good, but for *every* good. The resulting change in the good's quantity demanded is just part of the overall rearrangement of spending that strives to keep maximising total utility given the new price.

Consider how the increase in the price of pizza affects the demand curve for beer. As we increase the price of pizza from £1 to £3 per slice, consumption of beer rises from three to five slices. But the price of beer was unchanged. This means that the demand curve for beer must have *shifted* (which we explain in Chapter 8).

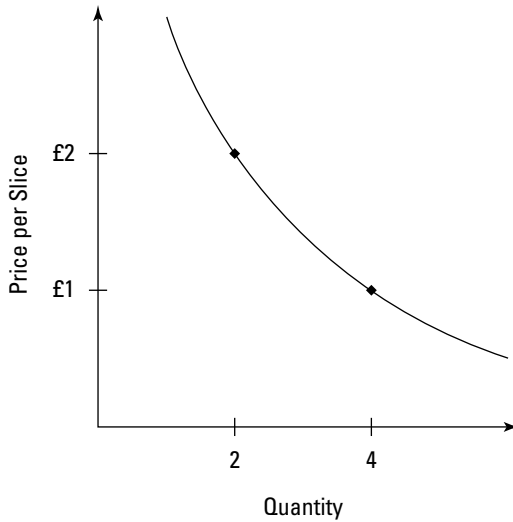


Figure 9-3:
The demand curve for slices of pizza.

We illustrate this shift in Figure 9-4. Point A on demand curve D shifts over to become point A' on demand curve D' . Events like this, where changes in the price of one good affect the quantity demanded of another good, are called *cross-price effects*. By contrast, when a change in a good's own price affects its own quantity demanded, you have *own-price effects*. Please note that whereas cross-price effects cause demand curves to shift, own-price effects cause movements along given demand curves.

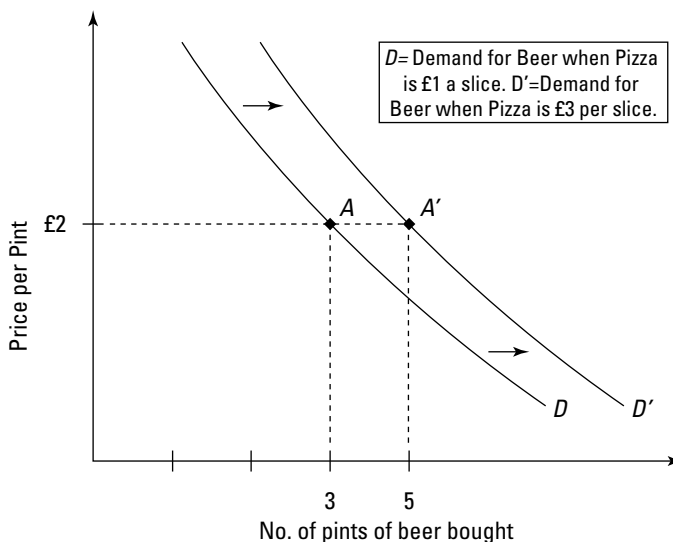


Figure 9-4:
The demand curve for beer shifts to the right when the price of pizza changes, holding the price of beer constant.

Complementary goods and substitute goods

Some things just go together: fish and chips; sausages and ketchup; shoes and shoelaces. In each of these pairs, the goods in question are more useful or more pleasing when consumed along with the other member of the pair.

Because such goods complement each other, economists refer to them as *complementary goods*. An interesting thing about complementary goods is that changes in the price of one complement affect the other complement. For example, the price of the petrol you put in your car affects the demand for cars; if the price of petrol goes up, fewer cars are sold.

In contrast, consider *substitute goods*, which are goods that serve similar functions so that if the price of one goes up, people switch to the other one. For example, if the price of train travel goes up, more people drive cars. And if

the cost of regular mail goes up, more people use email.

Both complementary goods and substitute goods are the result of cross-price effects. An increase in the price of a complement causes the quantity demanded of its pair to fall, whereas an increase in the price of a substitute causes the quantity demanded of its pair to rise.

As you look around the economy, make sure that you think of it as one great big organic whole, where things don't happen in isolation. When the price of one good changes, it doesn't affect just that good, but also many other goods that are substitutes or complements. And if the prices of the substitutes or complements change too, as a result of the initial price change, all their substitutes and complements are also affected, like a gigantic ripple effect.

The direction of a cross-price effect depends on the situation. In this chapter, we allow consumers to purchase only two goods, beer and pizza. The result is that when the price of pizza goes up, they switch some of their purchasing power over to buying beer – or, as economists say, they *substitute* from one good to the other (see the sidebar 'Complementary goods and substitute goods'). That's why when the price of pizza goes up, the demand curve for beer in Figure 9-4 shifts to the right.

But in the real world, where many other consumption goods are available, the demand curve may very well shift in the other direction. For example, some people like drinking beer only when they eat pizza. For them, an increase in the price of pizza may decrease *both* the amount of pizza eaten and the amount of beer drunk.

Such people think of beer and pizza as a bundle. An increase in the price of one member of the bundle increases the price of the entire bundle. These people buy less of each member of the bundle in order to free up money to spend on the many other consumption goods available. For consumers with these preferences and with the option of buying goods besides beer and pizza, when the price of pizza goes up, the demand curve for beer shifts left.

Chapter 10

The Core of Capitalism: The Profit-Maximising Firm

In This Chapter

- ▶ Understanding why firms choose to maximise profits
- ▶ Factoring competition into the mix
- ▶ Deconstructing a firm's cost structure
- ▶ Determining a firm's profit-maximising output level
- ▶ Seeing how costs determine a firm's supply curve
- ▶ Understanding how firms react to losing money

In modern market economies like the UK, some sort of business enterprise made nearly everything you eat, drink, wear, drive, ride, fly or use. So, naturally, economists devote a huge amount of effort to studying how businesses behave.

In this chapter, we show you how economists model a firm that's a member of a competitive industry; a firm that's just one of many firms competing against each other for your business. You need to understand how firms behave in competitive industries for two reasons:

- ✔ Most firms in the real world face a lot of competition because they are members of industries in which firms can enter and exit relatively freely. Therefore, firms have to worry about competitors already in the industry and those that may potentially come into the industry.
- ✔ All firms – even those that don't face much competition – behave in remarkably similar ways.

Above all, firms like to maximise profits. And, even more importantly, all firms go about maximising profits in the same way: by producing exactly the level of output at which the cost of producing one more unit just equals the increase in revenue that the firm gets from selling that unit.

In this chapter, we show you why firms behave in this way. Knowing the reasons helps you understand how *all* firms work, whether they face strong competition from rivals or have no rivals at all.

Maximising Profits Is a Firm's Goal

People create firms in order to produce things. That statement may prompt you to ask a fundamental question: *why* do people bother creating firms to make things? One reason may be altruism. Another may be that making things is fun. Another may be that the people who start a firm are bored doing other things. But economists think the answer is much simpler.



Economists assume that the overriding goal of all firms is to make as big a profit as possible. Economists make this assumption for two reasons:

- ✓ Every firm has profit maximisation near the top of its to-do list.
- ✓ Every firm wants to maximise profits after taking steps to achieve whatever other goals it may have.

For example, a firm may want to have a factory that emits no greenhouse gases. Yet, after it builds such a factory, the firm still wants to make as much money as possible. After all, after the necessary steps have been taken to protect the environment, why not make a nice big profit?



When the ice cream company Ben & Jerry's started, it donated a large percentage of its profits to charity. Given such a policy, the best way to help worthy causes was for Ben & Jerry's to make as big a profit as possible. In the UK, ethical cosmetics company The Body Shop and smoothie maker Innocent follow a social purpose and do so as profit-making companies.

To put it a bit more bluntly, if you don't make at least some profits in the long run, you aren't going to have a business for very long. Even if your personal goals for your business aren't about making yourself a nice big retirement fund, you're going to have to worry about making profits at some level or another.

Some non-economists object to people and businesses earning profits, but try thinking about it from the following perspective. If profits are measured by the difference between benefits (revenues) and costs, then profits accruing to your business are another way of saying that the people value whatever it is you're selling more highly than the cost to you of making it. Also, you can take your profits and instead of spending them in whichever high-rollers' hangout attracts you, you can reinvest them, creating more things of more value to more people, or to create jobs. The fact is that, like most things in economics, profit isn't a moral measure. You can use the excess

created in wiser or stupider ways. The existence of profit is merely a way of saying that at some level people valued your enterprise sufficiently for you to stay in business. (We'll come back to this point a bit later, when we discuss how market structure affects profitability.)

Facing Competition

Firms may or may not face a lot of competition from other firms. At one extreme lies *monopoly*, in which a firm is the only firm in its industry and faces no competition. At the other extreme lies what economists call *perfect competition*, a situation in which a firm competes against many other firms in an industry in which they all produce an identical good. And in between the extremes lie two situations: *oligopoly*, where two, three or (at most) a few firms are in an industry; and *imperfect (monopolistic) competition*, in which many competitors exist, but each produces a good that is unique – at least in some way – to the producing company. (See Chapters 12 and 13 for details on monopolies, oligopolies and monopolistic competition.)

In this chapter, you find out how firms behave under perfect competition. This situation is the simplest case to understand, because in an industry in which many competitors are producing identical products, none of the firms has any control over the price they charge.

Listing the requirements for perfect competition



To know why firms engaging in perfect competition have no control over the prices they charge, you have to understand that perfect competition assumes three things about the firms in an industry:

- ✓ Each firm is one of many in the industry.
- ✓ Each firm represents a very small part of the industry.
- ✓ Each firm sells identical or nearly identical products.



Once upon a time, people used to do their shopping in markets, and a market comes close to fulfilling the three criteria above. Individual stallholders are gathered together so that each has only a small share of the industry (in this case, a market). This situation means that individual stallholders are *price takers*, that is, no single stallholder can affect the price that consumers are willing to pay for their produce. A nice shiny apple is therefore likely to cost the same no matter which vendor you buy it from!

Commodity markets also have this property. A *commodity* is something that has a defined quality, such as a metal. For example, something either is gold or it is not gold. In the wholesale market, there's no such thing as being more 'goldy' than something else. The gold that's traded on commodity exchanges in London either satisfies this definition or it's something else.

To see why these things together mean that individual miners have no control over the price of gold, start with the fact that they are producing a nearly identical product. Because the gold from one place is identical to the gold from any other place, the only way an Angolan producer can entice you to buy from him rather than from a Russian producer is to offer you a lower price. Because all the gold is identical, all that producers have to compete on is price and price alone.

With price jumping to the fore as the key factor in the commodity market, we can use supply and demand analysis to figure out what the price is going to be. As we describe in Chapter 8, where the *market demand curve* for gold crosses the *market supply curve* for gold determines the price. How are these curves determined?

- ✔ To determine the *market demand curve* for gold, add up the individual demand curves of all the people who want to buy gold.
- ✔ To determine the *market supply curve* for gold, add up the individual supply curves of all the individual gold producers.

The first two assumptions of perfect competition come into play here: so many producers of gold exist and each producer produces such a very small part of the total supply of gold that the market supply curve for gold is basically unaffected by the presence or absence of any given individual supply curve of any particular gold producer. If a trillion troy ounces of gold are sold every year, the market price is unaffected by whether a small producer with only 1,000 ounces to sell bothers showing up to the market or not. He's just too small a player to cause the market price to change.

If every player is too small to cause the market price to change, each one has to take as given whatever price is generated by market demand interacting with market supply.

Acting as price takers but quantity makers

When the three assumptions of perfect competition are met, they produce a situation in which individual firms have no control over the prices they can charge. If that's all true, the individual firms can justly be called price takers.

Given that you have no overall control over production processes or how different you can make your product, even the most powerful firm can hope to control only two things: how much of its product to make and what price to charge. Because firms have no control over their prices under perfect competition, the list narrows to one: the only thing that price-taking firms can control is how much to produce.

Firms choose to make whatever quantity maximises their profits. This fact is mathematically convenient because it turns out that the quantity of output that a firm chooses to produce controls each of the two things that determine profits: total revenues and total costs.

To understand this fact more clearly, you have to know that a firm's profit is simply defined as its total revenue minus its total costs. Put into algebra,

$$\text{Profit} = TR - TC \quad (1)$$

where TR stands for total revenue, and TC stands for total costs.

The total revenue for a competitive firm is simply the quantity, q , of its output that it chooses to sell times the market price, p , that it can get for each unit:

$$TR = p \times q \quad (2)$$



For example, if a market vendor can sell 37 apples for £1 each, his total revenue is £37. (Yes, those apples are expensive, but you get the idea.) But notice that because the price at which the vendor can sell (p) is out of his hands if he's a price taker, the only way he can control total revenue is by deciding how many apples to sell. So a firm's decision about how big or small to make q can determine its total revenue.

Much of the rest of this chapter is devoted to showing you that the firm's total costs, TC , are also determined by how big or small q is. But, interestingly, although each extra unit of q sold brings in p pounds of revenue, the cost of each unit of q manufactured depends on how many units of q have already been made. Costs tend to increase as firms produce more and more, so each successive unit costs more than the previous unit. This fact ends up limiting the number of units that a firm wants to produce.

For example, suppose that the apple vendor can sell as many apples as he wants for £1 each. The first apple costs 10 pence to produce, the second one costs 20 pence, the third one costs 30 pence, and so on. In such a case, he's willing to produce no more than ten apples. Why? Because for each of the first nine apples, he makes a profit, but for apple ten (which costs £1 to

produce), he only breaks even. If he produces any more apples, he sustains a loss. (Apple number 11, for example, costs £1.10 to produce, but he gets only £1 for selling it.)

Consequently, you can see that the firm's choice of q determines both the TR and TC terms in profit equation (1). The only thing left to figure out is exactly how big to make q in order to maximise profits. Fortunately a ridiculously simple formula exists that gives the solution. Pay attention because you just may, uh, profit from reading this chapter. (Sorry!)

But before we get to the formula, we need to clarify a major source of confusion caused by the fact that economists use the word *profit* to mean something slightly different from conventional usage.

Distinguishing between accounting profits and economic profits

To an economist, the terms *profit* and *loss* refer to whether the gains from running a business are bigger or smaller than the costs involved. If the gains exceed the costs, you're said to be *running a profit*, whereas if the costs exceed the gains, you're said to be *running a loss*. If the two are just equal, you're said to be *breaking even*.

Things get complicated, however, because although accountants and economists agree on what counts as revenue, they disagree on what to count as costs.



Taking account of opportunity costs

Consider a business that sells lemonade. Both the accountant and the economist agree that the firm's revenues are simply how much money it makes from selling lemonade. However, they differ on what to count as costs:

- ✓ Accountants consider costs to be only actual monies spent in running the business: how much the firm pays its workers, how much it pays to buy lemons and so on. If the firm has revenues of £10,000, and it spends £9,000 to make those revenues, the accountant concludes that the firm's profit is £1,000. This number is the firm's *accounting profit* – the type of profit that is reported every day in financial statements and newspaper articles.
- ✓ Economists prefer a subtler concept, referred to as *economic profit*. Economic profit takes into account not just the money costs directly incurred by running a business, but also the *opportunity costs* incurred.

As we explain in Chapter 2, opportunity costs are what you have to give up in order to do something. Think about the entrepreneur who starts this lemonade business. After paying for materials and employees' wages, the accounting profits are £1,000. But is that really a good deal?

Suppose that this person left a job as a computer programmer to open up the lemonade business, and in the same amount of time that it took the lemonade business to turn a £1,000 profit, she would have made £10,000 in wages if she had stayed at her old job. That is, she gave up the opportunity to earn £10,000 in wages to open up a business that makes only a £1,000 accounting profit. She actually sustains an *economic loss* of £9,000. When you know this fact, her decision to switch careers doesn't seem like such a good idea.

Being motivated by economic profits

Economists like to concentrate on economic profits and losses rather than accounting profits or losses because the economic profits and losses are what motivate people. In our example, you can imagine that when other computer programmers see what happened to their former colleague after she switched careers, they're not going to follow her.

For the rest of the chapter, whenever you see any costs listed, assume that they are *economic costs*; that is, they include not only money directly spent operating a business, but also the costs of other opportunities foregone in order to operate the business. Likewise, whenever you see a profit or a loss, assume that this is an economic profit or an economic loss – the factor that motivates entrepreneurs to want to do something or to avoid doing it.

The most important application of this concept is to determine how much output a firm needs to produce. If producing the 12th unit of a product produces an economic profit, obviously the firm wants to produce it. But if increasing production to a 13th unit results in an economic loss, obviously the firm doesn't want to produce it.

By taking into account economic profits and losses, you get directly at what motivates firms to produce not only the types of goods they choose to produce, but also the quantities of those goods.

Analysing a Firm's Cost Structure

To understand how costs and revenues interact to determine economic profits or losses, economists like to break up a firm's total costs into two subcategories: fixed costs and variable costs.

- ✓ *Fixed costs* are costs that have to be paid even if the firm isn't producing anything. For example, after a rent contract is signed for the firm's headquarters, that rent must be paid whether the firm produces anything or not. Similarly, if the firm has taken out a loan, it's legally required to make its debt payments whether producing zero units of output or a billion units of output.
- ✓ *Variable costs* are costs that vary with the amount of output produced. For example, if you are in the lemonade-making business and you choose to produce nothing, you obviously don't have to buy any lemons. But the more lemonade you do produce, the more you spend buying lemons. Similarly, producing more lemonade requires more workers, so your labour costs also vary with the amount of output you produce.

Fixed costs can be represented as FC and variable costs as VC . Together, they sum up to a firm's total costs, or TC :

$$TC = FC + VC \quad (3)$$

As you look at equation (3), keep in mind that it deals with the economic costs facing the firm and therefore captures the opportunity costs of the firm's expenditures on both fixed costs and variable costs. (All expenditures, whether they're fixed costs or variable costs, involve opportunity costs – the other things you gave up buying in order to spend the money you spent on your fixed and variable costs.)

Focusing on costs per unit of output

Economists distinguish between fixed and variable costs because they have very different effects on a firm's decision of how much to produce. Take a look at Table 10-1, which gives data on LemonAid Ltd, our lemonade producer. To keep things simple, we're varying only the amount of labour involved in producing a bottle.



When LemonAid gets started, it buys a juicer machine for £100, which gives it fixed costs of £100. It then has to decide how much to produce, which in turn determines how many workers it needs to hire. In the first column, the number of workers varies from zero to eight. If the firm hires no workers, you can see in the top entry of the second column that no output is produced. But if it hires workers, output increases as you move down the second column. More workers mean more output.

<i>Workers</i>	<i>Output</i>	<i>Fixed Costs</i>	<i>Average Fixed Costs</i>	<i>Variable Costs</i>	<i>Average Variable Costs</i>	<i>Total Costs</i>	<i>Average Total Costs</i>	<i>Marginal Costs</i>
0	0	100	–	0	–	100	–	–
1	50	100	2.00	80	1.60	180	3.60	1.60
2	140	100	0.71	160	1.14	260	1.86	0.89
3	220	100	0.45	240	1.09	340	1.55	1.00
4	290	100	0.34	320	1.10	420	1.45	1.14
5	350	100	0.29	400	1.14	500	1.43	1.33
6	400	100	0.25	480	1.20	580	1.45	1.60
7	440	100	0.23	560	1.27	660	1.50	2.00
8	470	100	0.21	640	1.36	740	1.57	2.67

Studying increasing and decreasing returns

Pay attention to the fact that the amount of additional, or marginal, output produced by each additional worker is not constant: that is, if you go from no workers to one worker, output increases from nothing to 50 bottles of lemonade. However, as you go from one worker to two workers, output increases from 50 bottles to 140 bottles. Put into economic jargon, the second worker's *marginal output* is 90 bottles, whereas the first worker's marginal output is only 50 bottles.

Now look at these facts in terms of costs and benefits. If you have to pay each worker the same wage of £80 per day (£10 per hour for 8 hours of work), you're going to like the fact that whereas the first worker produces 50 bottles for his £80 pay, the second worker produces 90 bottles for her £80 pay.

Economists refer to situations like this as *increasing returns*, because the amount of return you get for a given amount of input (one more worker) increases as you add successive units of input. But if you look farther down the second column, you find that increasing returns don't last forever.

Indeed, in the case of LemonAid, increasing returns end almost immediately. Consider what happens to output when you add a third worker. Output does increase, but only by 80 units, from 140 bottles to 220 bottles. And things get even worse the more workers you add. Adding a fourth worker increases output by only 70 bottles, and adding a fifth increases output by only 60 bottles.

Economists call situations like this *diminishing returns*, because each successive unit of an input, such as labour, brings with it a smaller increase in output than the previous unit of input.

Determining the cause of diminishing returns

We go into more detail about what causes diminishing returns in Chapter 3, but we explain the process briefly here. What's going on is that LemonAid bought only one juicer machine for squeezing the juice out of lemons.

The first worker can use the machine to squeeze enough juice for 50 bottles by carrying lemons to the machine and then operating the machine. But it turns out that two workers together can do even better by dividing up the work: one brings lemons to the machine, and the other operates it. Working together, they can produce a total of 140 bottles – more than double the 50 bottles that one worker can produce working alone.

However, a third worker doesn't increase output nearly as much as a second because the two major tasks – carrying and operating – have already been taken care of. At best, he can just help the first two workers do these tasks a little faster. The same holds true for all successive workers: having them is helpful, but each one adds less to output than the previous one because things start getting crowded and little room is left for improvement.

Examining average variable costs

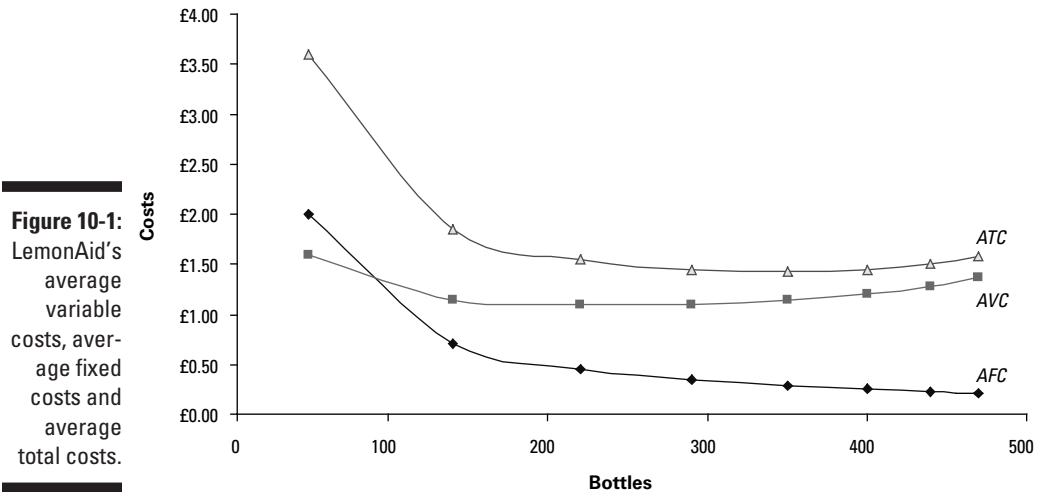
Variable costs are affected by the fact that additional workers first bring increasing returns but then decreasing returns. In the case of the LemonAid example in Table 10-1, the variable costs are all labour costs, with each worker having to be paid £80 per day. You can see these variable costs increase as you move down the fifth column.

But what's much more interesting is looking at *average variable costs (AVC)*, which are defined as variable costs divided by quantity (VC/q). For example, because one worker produces 50 bottles of output at a variable cost of £80, the average variable cost is $£80/50 = £1.60$ per bottle. When two workers together cost £160 in variable costs but produce 140 bottles, the average variable cost for two workers is only $£160/£140 = £1.14$ per bottle.

The decrease in average variable costs is the result of increasing returns: when moving from one worker to two workers, variable costs double (from £80 to £160) but output more than doubles (from 50 bottles to 140 bottles).

When diminishing returns set in, average variable costs start to rise, which you can see as you move down the sixth column of Table 10-1. This happens because although each additional worker costs an extra £80, each additional worker after the second worker brings a smaller increase in output than his predecessor. Each successive £80 wage payment brings with it fewer and fewer additional bottles produced, so the average variable cost per bottle must rise.

LemonAid's average variable costs show up as a subtle U shape when you plot them on a graph, which we do in Figure 10-1. (We also show the company's average fixed costs and average total costs.) Keep this average variable cost curve in mind because it's going to have a huge effect on how many bottles the firm's managers want to produce in order to maximise firm profits.



Watching average fixed costs fall

Average fixed costs (AFC) are defined as fixed costs divided by quantity (FC/q). The fixed costs of LemonAid are always the £100 the firm paid for the juicer machine, no matter what amount of output it produces. As a result, the more lemonade it produces, the lower its average fixed costs. That's why AFC falls (see the fourth column of Table 10-1) from a value of £2.00 per bottle when 50 bottles are produced using one worker down to only £0.21 per bottle when 470 bottles are produced using eight workers.



Average fixed costs always decline, because the same fixed cost gets divided up over a greater and greater number of units of output as output increases. When you plot out average fixed costs per bottle, as in Figure 10-1, you get a downward sloping AFC curve. Keep this fact in mind because it helps explain the shape of the average total costs (ATC) curve, as we explain in the next section.

Tracking the movement of average total costs

In the previous two sections, we show you that average fixed costs always decline as output increases, whereas average variable costs first fall (due to increasing returns) and then rise (due to diminishing returns). Because total costs are the sum of fixed costs and variable costs, *average total costs* obviously depend on how average fixed costs and average variable costs sum up.

Average total costs (*ATC*) are defined as total costs divided by quantity (TC/q). Now, take a look back at equation (3) earlier in the chapter. If you divide every term in equation (3) by q , you get the following:

$$TC/q = FC/q + VC/q \quad (4)$$

You can simplify equation (4) by realising that $ATC = TC/q$, $AFC = FC/q$, and $AVC = VC/q$. What you get is

$$ATC = AFC + AVC \quad (5)$$

You can see clearly from equation (5) that average total costs depend on how average fixed costs and average variable costs interact. You need to understand two key points here:

- ✓ *ATC* must always be greater than *AVC*, because you have to add in *AFC*.
- ✓ *ATC* reaches its minimum value at a higher level of output than *AVC*.

To see that the first point is true, look at Figure 10-1, which shows that the *ATC* curve is above the *AVC* curve. The vertical distance between them at any particular level of output is equal to the *AFC* at that output level. As you move from lower output levels to higher output levels, the *ATC* and *AVC* curves converge because *AFC* becomes smaller and smaller. (In other words, the vertical distance between the *ATC* and *AVC* curves also gets smaller and smaller.)

To see that the second point is true, look at Table 10-1 again. You can see that average variable costs reach their minimum value of £1.09, when three workers are hired and 220 bottles are produced. Average total costs, however, reach their minimum of £1.43 when five workers are hired and 350 bottles are produced.

This happens because average fixed costs are always falling, meaning that in equation (5), the *AFC* part on the right-hand side of the equation is always getting smaller and smaller. This constant decline helps to offset temporarily the increases in average variable costs that happen when diminishing returns set in. Consequently, although average variable costs bottom out at three workers, average total costs don't bottom out and start increasing until the fifth worker.

Focusing on marginal costs

The manager of a firm wants to know what quantity, q , of output she needs to produce in order to maximise profits. To solve this problem, she needs one more cost concept: marginal cost.

Marginal cost is how much total costs increase when you produce one more unit of output. The marginal cost of one more unit of output depends on how much output has already been produced.

To see this, examine the total costs column of Table 10-1. Notice that total costs increase from £100 in the first row to £180 in the second row as output increases from 0 bottles to 50 bottles when the firm hires the first worker. In other words, costs go up £80 while output goes up 50 bottles. So each of these extra, *marginal* 50 bottles on average increases costs by $£80/50 = £1.60$ each. The marginal cost per bottle, *MC*, is defined as follows:

$$MC = (\text{Change in } TC)/(\text{Change in } q) \quad (6)$$

As you move down the marginal costs column of Table 10-1, you can see that marginal costs first fall and then rise: yet another reflection of the fact that LemonAid's production process exhibits increasing returns followed by diminishing returns. Because the second worker produces much more than the first worker but costs the same, the marginal cost falls when the second worker is added. For successive workers, costs keep increasing but marginal output keeps declining, which means marginal costs must rise.

Noticing where the MC curve crosses the AVC and ATC curves

Here's a fun fact that economists love: if you plot out marginal costs to create a marginal cost (*MC*) curve, that curve crosses both the average variable cost (*AVC*) curve and average total cost (*ATC*) curve at their minimum points – that is, at the bottom of their respective U shapes. (What, you don't see the cause for celebration? Keep reading – maybe we can boost your enjoyment level.)

In Figure 10-2, we plot the *AVC*, *ATC* and *MC* curves that you get by plotting out the data in Table 10-1. The *MC* curve goes through the minimum points of both the *AVC* and *ATC* curves.

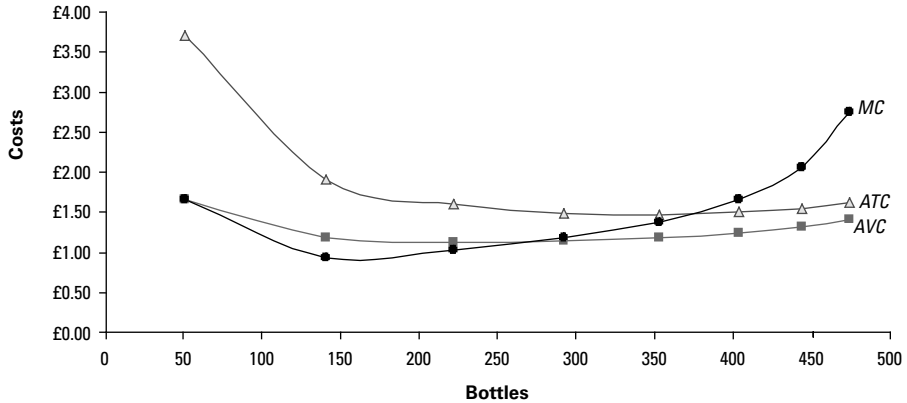


Figure 10-2:
LemonAid's
AVC, *ATC*,
and *MC*
curves.

This happens because the marginal cost at each unit determines whether the *AVC* and *ATC* curves are increasing or decreasing. Huh? To simplify, we change our example for a moment; instead of thinking about costs, we think about heights.



Think about a room with ten people in it. Suppose you determine that the average height of the people in the room is 156 centimetres. Now think about what happens to that average when another person walks into the room:

- ✓ If the 11th person is taller than the previous average, the average rises.
- ✓ If the 11th person is shorter than average, the average falls.
- ✓ If the 11th person is exactly 156 centimetres tall, the average stays the same.

The same sort of reasoning applies to marginal costs and average costs. After q units of output, you can figure out *AVC* and *ATC*, just like you can compute the average height after the first ten people enter the room. After that, *AVC* and *ATC* rise or fall depending on the *MC* of the next unit of output, just as the average height of the people in the room increases, decreases or stays the same, depending on the height of the next person entering the room. Here's what we mean:

- ✓ If the *MC* is less than the previous average costs, the averages fall.
- ✓ If the *MC* is greater than the previous average costs, the averages rise.
- ✓ If the *MC* is exactly the same as the previous average costs, the averages stay the same.

You can see these effects graphically by looking at various parts of Figure 10-2. First, look at the output level of 140 bottles (why not 150? We come to that in the next section). At that output level, the *MC* of producing one more bottle is less than both *ATC* and *AVC*, meaning that *ATC* and *AVC* decreases if output is increased by one more bottle. That's why the *AVC* curve and the *ATC* curve are downward sloping at that output level. The average curves are being pulled down by the low value of *MC*.

Next, look at the output level of 440 bottles. You can see that the *MC* at that output level is higher than the *ATC* and the *AVC*. Consequently, both *AVC* and *ATC* must be increasing, as reflected geometrically by the upward slopes of both the *AVC* curve and the *ATC* curve. The curves slope upwards because the high value for *MC* is pulling them up.

Now, to put some pieces together, notice that the *MC* curve causes both the *AVC* curve and the *ATC* curve to be U-shaped (albeit subtly). On the left side of Figure 10-2, the fact that *MC* is less than the average curves means that the average curves slope downward. On the right side of Figure 10-2, the fact that *MC* is greater than the average curves means that the average curves slope upward.

So we've come full circle to the fact that the *MC* curve has to cross the two average curves at their respective minimum points – at the bottoms of their respective U shapes. To the left of such a crossing point, the average must be falling because *MC* is less than the average. And to the right, the average must be rising because *MC* is larger than the average. But where the curves cross, the average curve is neither rising nor falling because the *MC* of that unit of output is equal to the current average. (In other words, a person who is 156 centimetres tall has walked into a room that already has an average height of 156 centimetres, so the average doesn't budge.)

Economists love to go on and on about this fact, but really it just reflects the effect that increasing and then decreasing returns have on cost curves. Costs first fall and then rise. And at some point in the middle, there's a magic level of output where increasing and decreasing returns are balanced, while transitioning from falling to rising. That point must be where marginal cost equals average cost, because only when *MC* equals average cost can average cost be stationary.

Comparing Marginal Revenues with Marginal Costs

In the previous section, we explain how marginal costs relate to average costs. With that info in mind, we're finally ready to explain how managers decide how much output to produce in order to maximise profits. (You thought we'd never get here, didn't you!)



Here's a sad but true fact to keep in mind: firms can't always make a profit. In a perfectly competitive industry, a firm can't control the price for which its output sells, and sometimes that price is too low for the firm to make a profit, whatever quantity it produces. When that happens, the best the firm can do is minimise its losses and hope for the price to change. If the price drops low enough, the best thing to do may be to shut down production immediately, because that way the firm loses only its fixed costs. (We discuss the difference between fixed and variable costs in the section 'Analysing a Firm's Cost Structure', earlier in the chapter.)

Later in the chapter (in the section, 'Pulling the Plug: When Producing Nothing Is Your Best Bet'), we discuss this sad situation in more detail. But first, we focus on a happier situation – one in which the market price is high enough that a firm wants to produce a positive amount of output. As you're going to see, this may or may not mean that a firm is making a profit, but even if it isn't, the losses aren't great enough to halt production.

The magic formula: Finding where $MR = MC$

In the typical case where market prices are high enough that a firm wants to make a positive amount of output, a ridiculously simple formula is used to determine the optimal quantity of output, q , that the firm needs to produce. The firm wants to produce at the level of output where marginal revenue equals marginal cost ($MR = MC$).

Producing where $MR = MC$ does two things:

- ✓ It minimises the firm's loss if it has to take a loss due to a low selling price for its output.
- ✓ It maximises the firm's profit if it can make a profit, because the selling price is high enough.

The idea behind $MR = MC$ is very simple and basically comes down to a cost versus benefit analysis. If producing and selling a bottle brings in more revenue than it costs to produce the bottle, produce it. If not, don't produce it. Easy, right?



Think back to our example again. Imagine that LemonAid can sell each bottle of lemonade that it produces for £2 each. Economists like to say that the *marginal revenue* of each bottle is £2, because each and every bottle when sold brings in an extra £2.

What the firm’s managers must do is decide how much to produce based on whether any given bottle costs more or less than the £2 marginal revenue that the firm would get by selling it.

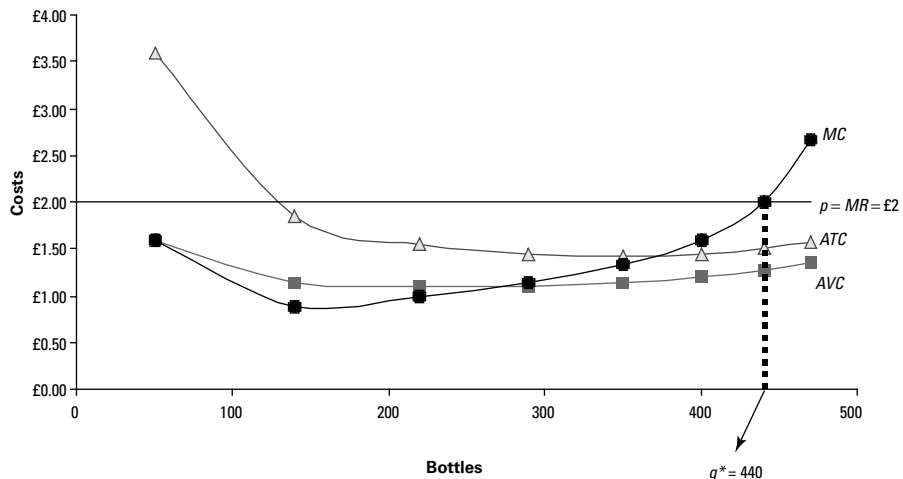
Be very careful at this point. You have to remember that the relevant cost that the managers look at is an individual bottle’s marginal cost, *MC*. That’s because if they’re deciding whether to produce that particular bottle, they need to isolate that bottle’s production cost from the costs of all previously produced bottles in order to compare it to the revenue that the bottle brings if produced and sold. *MC* does just that by ignoring all previous bottles and focusing on what the next bottle is going to cost to make.

If the *MC* of that bottle is less than £2, obviously producing that bottle makes a gain, and so the managers choose to produce it. On the other hand, if the *MC* is bigger than £2, producing the bottle causes a loss, and the managers choose not to produce it.

By looking at the *MC* of every possible bottle (the 1st, the 5th, the 97th and so on) and comparing it with marginal revenue that the firm can get by selling it, the managers can determine exactly how many bottles to produce. The necessary comparisons can be done by looking at a table of costs, such as Table 10-1, but making the comparisons graphically is even easier.

In Figure 10-3, we draw in the marginal cost (*MC*), average variable cost (*AVC*) and average total cost (*ATC*) curves for LemonAid. We also draw in a horizontal line at £2, which is the marginal revenue for selling any and all bottles that the firm may choose to produce. We label the line $p = MR = £2$ to indicate the fact that the selling price of the bottle is £2, which is also the marginal revenue.

Figure 10-3:
The firm’s optimal output level, q^* , happens where *MC* and *MR* cross.



The quantity q^* , which corresponds to where the horizontal $p = MR = £2$ line crosses the MC curve, shows that $q^* = 440$ bottles. This quantity is the level of output that the firm chooses to produce in order to maximise profits.

To understand why adhering to $MR = MC$ maximises profits, look back at Table 10-1 earlier in the chapter and consider each unit of output, q , for which $q < 440$. For all these units, the marginal revenue is greater than the marginal cost ($MR > MC$), meaning that producing and selling each of these bottles brings in more money than it costs to make them. For example, look at bottle number 140. It has a marginal cost of only £0.89 but can be sold for £2. Clearly, you need to make such a bottle because you make more selling it than it costs to produce. The same is true for all the bottles for which $q < 440$; you need to produce them all because they all bring in a profit.

On the other hand, for all units above the q^* level of output ($q > 440$), the case is reversed: the marginal revenue is less than the marginal cost ($MR < MC$). You lose money if you produce and sell those bottles. For example, at an output level of 470 bottles, the MC is £2.67 while the MR is only £2. If you produce at that output level, you lose 67 pence on bottle number 470. Clearly, you don't want to do this.

By comparing the marginal revenues and marginal costs at all output levels, you can see that the managers of LemonAid want to produce exactly $q^* = 440$ units, the number of units where the MR and MC lines cross.

As we mention in the introduction to this section, producing where $MR = MC$ doesn't guarantee you a profit, but it does at least make sure that you only produce bottles that bring in more money than they cost to make. This formula by itself can't guarantee a profit because it doesn't take account of the fixed costs you have to pay no matter what level of output you're producing. Even though you only produce bottles for which marginal revenue is at least as great as marginal cost, you still may not make enough of a gain from these bottles to pay off your fixed costs.

Visualising profits

Here's what we know from the previous section:

- ✓ Producing when $MR = MC$ allows a firm to determine its optimal (best possible) output level, q^* .
- ✓ Producing at q^* doesn't guarantee a profit – instead, it guarantees that you're making the biggest profit possible (if a profit is possible) or the smallest loss possible (if prices are so low that no way to make a profit exists given your cost structure).

We're now going to show you a quick and easy way to use the cost curves visually to determine whether the firm is making a profit or a loss. The trick is to realise that the two components of profits, total revenue (TR) and total costs (TC), can each be represented by rectangles whose areas are equivalent to their respective sizes. As a result, you can immediately tell if profits are positive or negative by determining whether the TR rectangle is larger or smaller than the TC rectangle. If the TR rectangle exceeds the size of the TC rectangle, profits are positive. And if the TR rectangle is smaller than the TC rectangle, profits are negative and the firm is running at a loss.

To see how this works, look at Figure 10-4, where we draw a generalised set of average total cost (ATC), average variable cost (AVC) and marginal cost (MC) curves, in addition to a horizontal line labelled $p = MR$ to indicate that price equals marginal revenue for this competitive firm. By *generalised*, we mean a typical-looking set of curves; we're no longer using the particular curves you get by plotting out LemonAid's costs. Switching to this generalised set of curves convinces you (we hope!) that the geometric way of determining the size of a firm's profits holds true for *any* set of cost curves.

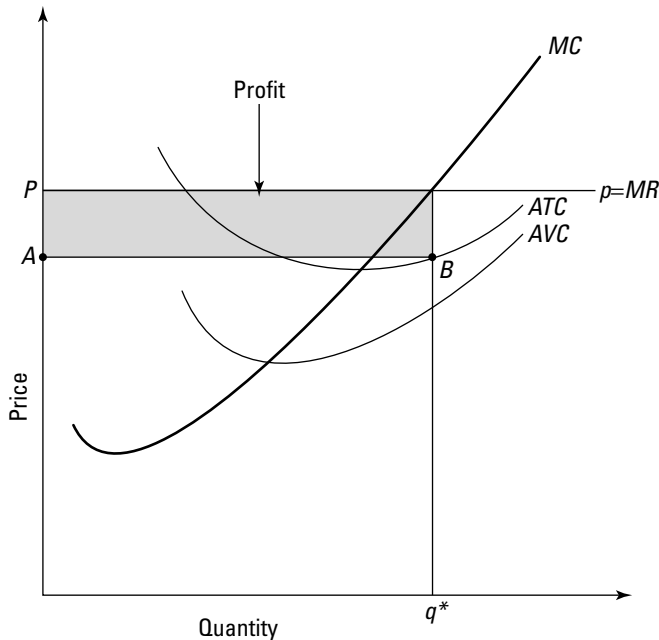


Figure 10-4:
A firm making a profit.



The big trick behind expressing total revenue as a rectangular area is to remember that, when producing the profit-maximising output level, q^* , a firm's total revenue is simply price times that quantity, or $TR = P \times q^*$. Just as you can define the area of a rectangular room as length times width, you can

define total revenue on a graph as a rectangle determined by price times quantity. In Figure 10-4, TR is a rectangle of height P and width q^* , with the four corners located at the origin, P , the point where the $p = MR$ line crosses the MC curve, and q^* .

You can also use a rectangle to represent the total costs that the firm incurs when producing q^* units of output. To figure out where to draw this rectangle, you have to use a little maths trick to convert the information that the average total cost (ATC) curve gives you into what you want to graph, total costs (TC).

To see how to apply this mathematical sleight of hand, first look at point B in Figure 10-4. B shows the average total cost (ATC) *per unit* when the firm is producing output level q^* . The trick is handy because it can be used to convince you that the rectangle whose width is q^* and whose height is given by the ATC at output level q^* is actually equal to the firm's total costs. That is, TC is equal to the area of the rectangle whose four corners are the origin, the point we label A , the point we label B , and q^* .

The heart of the maths trick is realising that when the firm is producing at q^* , $ATC = TC/q^*$. If you multiply both sides of this equation by q^* , you find that $ATC \times q^* = TC$. This equation tells you that TC is indeed equal to the product of ATC and q^* , or to the area of a rectangle of height ATC and width q^* – exactly the rectangle that we just showed you!

Now that you understand how the areas of rectangles that are derived from the firm's cost curves can represent a firm's TR and TC , you won't be surprised to discover that the firm's profits, which are by definition equal to $TR - TC$, can also be represented by the area of a specific rectangle. In fact, the profit is equal to the area of the shaded rectangle in Figure 10-4, because profits are simply the difference between TR and TC . Because the TR rectangle is larger than the TC rectangle in this case, the firm is making a profit whose size is equivalent to the area of the shaded rectangle that's defined by the area of the larger TR rectangle minus the area of the smaller TC rectangle.

An informative thing to do is to run a thought experiment using Figure 10-4. Imagine what happens if the price, P , increases. First, notice that the optimal output, q^* , increases, because the place where the horizontal $p = MR$ line crosses the MC curve moves up and to the right. Simultaneously, the total revenue rectangle increases in size, as does the total cost rectangle. But which one grows faster? Do profits rise or fall?

Go ahead and draw in some lines to convince yourself that profits do in fact increase – that is, the shaded profit rectangle grows in size as the price increases. As you discover, a rising price increases the firm's profits. The next section explains how profits can go negative if the price falls far enough.

Visualising losses

Compare the situation in the previous section to the one illustrated in Figure 10-5, where the cost curves are the same as in Figure 10-4 but the price (and therefore the marginal revenue [MR]) at which the firm can sell its product is much lower.

Following the $MR = MC$ rule for selecting the optimal output level, the firm chooses to produce at the output level q_2^* where the new lower $p = MR$ line crosses the MC curve. But the low price at which the firm is forced to sell its output means it can't make a profit. (We label the optimal output level for the firm in Figure 10-5 q_2^* to make clear that the optimal output level in this case, where the price is lower, is different from the optimal output level q^* in Figure 10-4 where the price was higher.)

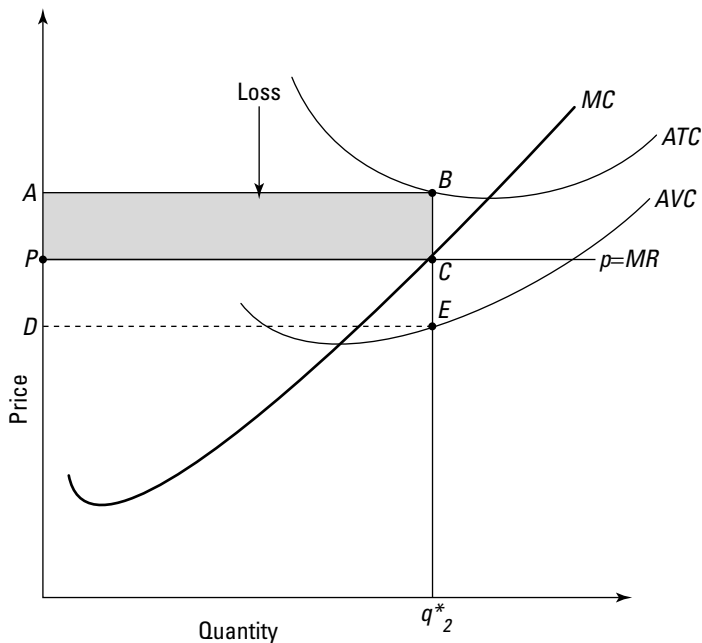


Figure 10-5:
A firm running a loss.

You can see the size of the loss geometrically by comparing the TR and the TC rectangles that occur in this situation. Because $TR = P \times q_2^*$, total revenue is equal to the area of a rectangle of height P and width q_2^* . Consequently, the TR is equal to the area of the rectangle whose four corners lie at the origin, P , C and q_2^* . The TR is smaller than the TC rectangle defined by the origin, point A , point B and q_2^* . Because the area of the total cost rectangle exceeds the area of the total revenue rectangle, the firm is running a loss equivalent to the size of the shaded area in Figure 10-5.

Figure 10-5 shows you that although a manager always wants to produce the level of output where $MR = MC$, doing so doesn't necessarily guarantee a profit. The problem is that fixed costs get in the way. For example, suppose a firm has to pay £1,000 a month in rent. If the month has already started and the rent has already been paid, you produce all units for which $MR > MC$. That gets you to output level q_2^* in Figure 10-5.

Suppose that $q_2^* = 600$ and the price at which you can sell output is £1 each. That makes for £600 in total revenue. But with £1,000 in rent costs, you still sustain a loss for the month even though the marginal revenue exceeds the marginal cost for each of the 600 units. The tricky part is that although marginal costs don't take fixed costs into account, profits do.

We say it again: Producing at the output level where $MR = MC$ doesn't guarantee a profit. But it does guarantee that if you have to run a loss, that loss is as small as possible. Although you can't do anything immediate about your fixed costs, you can make sure to produce only those units for which the marginal revenue from selling them is larger than the marginal cost of producing them.

Pulling the Plug: When Producing Nothing Is Your Best Bet

You may wonder why a firm stays in business when running a loss rather than a profit. The usual answer is that it hopes things are going to turn around soon: the firm expects the price at which it can sell its products to rise, or it expects that it can somehow reduce its costs of production.

Even if these expectations are well founded, a firm may still be better off completely shutting down production rather than producing some positive amount of output. The determining factor is once again fixed costs.

The short-run shutdown condition: Variable costs exceed total revenues

Suppose you're in charge of a firm with a monthly rent of £1,000. If you produce nothing, you sustain a loss of £1,000. But that doesn't mean you *definitely* start producing stuff in order to try to make back some of that money. Instead, you want to produce only if by doing so you are better off than if you do nothing. That is, you choose to produce if doing so results in an outright profit or a loss of less than the £1,000 you stand to lose by doing nothing. As we're about to show you, sometimes the best thing to do is to produce nothing.

Consider Figure 10-6, where the price at which the firm can sell its output is so low that the marginal revenue ($p = MR$) line and the marginal cost (MC) curve intersect at a point *below* the average variable cost (AVC) curve. What does this mean? Put simply, the total revenues in this case are actually *less than* variable costs. (Total revenues are represented by the rectangle whose four corners are at the origin and points P , B and Q^* , where Q^* represents the optimal output level at this price. Variable costs are represented by the rectangle whose four corners are the origin and points C , D and Q^* .)

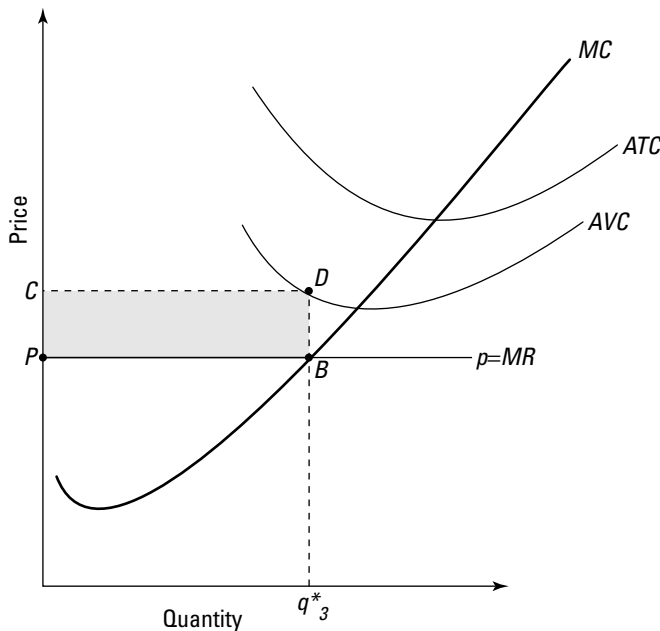


Figure 10-6:
A firm running such a huge loss that it can't even cover its variable costs.

This situation means that by producing Q^* units, the firm doesn't even bring in enough total revenue to cover the variable costs associated with producing that many units. The firm is not only going to lose its fixed costs, but also even more money by not being able to cover the variable costs associated with producing Q^* .

The logical thing to do in such a situation is to produce nothing. By producing zero units, you lose only your fixed costs. By producing Q^* , you lose even more money because you can't even cover your variable costs.

For example, suppose that fixed costs are £1,000 and that by producing Q^* units the firm makes total revenues of £400 and incurs variable costs of £500. Because total revenues cover only £400 of the £500 in variable costs, the firm loses £100 in variable costs by producing. Add to that the £1,000 of fixed

costs it incurs no matter how much it produces, and the firm loses a total of £1,100 by producing Q^* units of output. In contrast, if the firm shuts down and produces nothing, it loses only the £1,000 in fixed costs. Clearly, in such a situation, the firm chooses to shut down.

Economists call this situation the *short-run shutdown condition*. If a firm's total revenues at Q^* are less than variable costs, better to shut down completely. Graphically, this happens any time the horizontal $p = MR$ line intersects the MC curve at a point below the U-shaped AVC curve. In all such situations, total revenues are less than variable costs, implying that shutting down is better than producing.



For the same reason, producing when you're losing money on each item you produce is never a good idea. Factory Records famously discovered this on the release of New Order's 'Blue Monday' single. The packaging turned out to be so expensive that every single copy sold lost the label money. (Worse still, the record was so successful that the losses per sale totalled up to a mind-boggling sum of money.) The packaging was, of course, later changed. Factory later, famously, ended up going bust, and the owners remembered as artists rather than businesspeople.

The long-run shutdown condition: Total costs exceed total revenues

In contrast, look back at Figure 10-5. In this case, the firm is more than covering its variable costs because total revenues (represented by the box whose four corners are the origin, and points P , C and q_2^*) exceed variable costs (represented by the box whose four corners are the origin and points D , E and q_2^*). Although this firm is losing money, it's better off producing q_2^* rather than $q = 0$ because total revenues exceed variable costs. The firm can take the extra money left over after paying variable costs and use it to pay off some of its fixed costs.



Suppose that its fixed costs are £1,000 and that when producing output level q_2^* the firm has a total revenue of £800 and variable costs of £700. The first £700 of the £800 in total revenues can go to paying off the variable costs, leaving £100 to pay off a portion of the £1,000 in fixed costs. The result is an overall loss of £900 rather than a £1,000 loss if it produces nothing.

A firm in the situation of Figure 10-5 continues to operate in the short run because by doing so it's better off than shutting down immediately. But the firm's still losing money. So although producing output in the short run is better, the firm eventually wants to stop losing money by closing down. As soon as its fixed cost contracts expire, it shuts down permanently.

At the mercy of the market price

Because competitive firms have to take the market price as given, the decision whether to continue operating is in some senses totally out of their hands. Only two possibilities exist:

- ✓ If the price is high enough, the firm makes a profit and stays in business in order to keep collecting the profit. Graphically, this happens whenever the horizontal $p = MR$ line crosses the MC curve at a point above the bottom of the U-shaped ATC curve, as in Figure 10-4.
- ✓ If the horizontal $p = MR$ line crosses the MC curve at a point below the bottom of the U-shaped ATC curve, the firm is taking a loss. What it does in such situations depends on how low the price is and, consequently, how big the loss is. Two possibilities (or conditions) exist, as we explain in the previous sections:

- The *short-run shutdown condition* occurs when a firm's total revenues are less than its variable costs. Graphically, this happens when the horizontal $p = MR$ line intersects the MC curve at a point below the low point of the U-shaped AVC curve, as in Figure 10-6.

In such a situation, the firm is better off shutting down immediately and losing only its fixed costs. Producing output in such a situation results in an even bigger loss.

- The *long-run shutdown condition* occurs when a firm's total revenues exceed its variable costs but are less than its total costs. Graphically, this happens in any situation where the horizontal $p = MR$ line intersects the MC curve at any point on the segment of the MC curve that lies above the bottom of the U-shaped AVC curve but below the bottom of the U-shaped ATC curve, as in Figure 10-5.

In such a situation, the firm is guaranteed to lose money. But as long as the firm is stuck with its current set of fixed-cost commitments, producing rather than shutting down immediately is preferable. If it produces, its total revenue exceeds its variable costs, meaning that it can use the excess to pay off at least part of its fixed costs. On the other hand, if it shuts down and produces nothing, it loses all its fixed costs and thereby does worse.

As you can see, the perfectly competitive firm is in some sense totally at the mercy of the market price. If the price is high, it makes profits. If the price is low, it sustains losses. And even then, its decision about whether to shut down immediately or keep operating at a loss until it can get out of its fixed cost commitments depends entirely on the price. Perfectly competitive firms have no control.

In the next chapter, we discuss non-competitive firms and how they have control over their market prices. As you can already gather from the dependency of competitive firms on the market price, having such control puts them in a far less precarious position.

Chapter 11

Why Economists Love Free Markets and Competition

In This Chapter

- ▶ Measuring the social benefits of different output levels
- ▶ Demonstrating that free markets maximise total surplus
- ▶ Reducing total surplus with taxes and price controls
- ▶ Producing at the lowest possible cost to society
- ▶ Adjusting to changes in supply and demand

Economists love competitive free markets – markets in which numerous buyers freely interact with numerous competitive firms. Indeed, economists firmly believe that when they work properly, competitive free markets are the very best way to convert society’s limited resources into the goods and services that people want to buy.

Why do economists place such great confidence in competitive free markets? Because the interaction of supply and demand (which we discuss in Chapter 8) leads to an outcome in which every produced unit of output satisfies two excellent conditions:

- ✔ Each unit is produced at the minimum cost possible, which means that no waste or inefficiency is involved.
- ✔ Each unit’s benefits exceed its costs: only output that makes the world better off gets produced.

Economists also love competitive free markets because they provide a gold standard against which all other economic institutions can be judged. In fact, economists refer to many economic problems as *market failures* precisely because they are instances in which, if markets were able to function properly, the problems would quickly go away.

In this chapter, we show you that competitive free markets ensure that benefits exceed costs for all the output produced. We also show you that competitive free markets produce the *socially optimal quantity* of output – the level that maximises the benefits that society can get from its limited supply of resources. Finally, we show you how competitive industries adjust to changes in supply and demand to ensure that everything that’s being produced is produced at the lowest possible cost to society.

The Beauty of Competitive Free Markets: Ensuring That Benefits Exceed Costs

Society has a limited amount of land, labour and capital out of which to make things. Consequently, society must be very attentive when working out how to best convert its limited resources into the goods and services that people most greatly desire.



Economists love competitive free markets because, *if they are operating properly*, these markets make sure that resources are allocated *optimally*. In particular, such markets ensure that resources go towards producing only output for which the benefits exceed the costs.

This point can be easily demonstrated using nothing more complicated than a supply and demand graph such as the type we introduce in Chapter 8. But before we show you how that’s done, we need to explain the conditions under which competitive free markets can function properly and thereby deliver such good results. (Please note that for brevity, we sometimes just refer to ‘free markets’ or ‘markets’ in this chapter instead of writing out ‘competitive free markets’ each time. We’re trying to maximise our resources here.)

Examining prerequisites for properly functioning markets



Free markets guarantee optimal outcomes only if the following conditions are met:

1. Buyers and sellers all have access to the same full and complete information about the good or service in question.
2. Property rights are set up so that the only way buyers can get the good or service in question is by paying sellers for it.
3. Supply curves capture all the production costs that firms incur in making the good or service in question.

4. Demand curves capture all the benefits that people derive from the good or service in question.
5. Numerous buyers and numerous sellers exist, so that nobody is big enough to affect the market price. This condition is often called the *price-taking assumption*, because everybody has to take prices as given.
6. The market price is completely free to adjust to equalise supply and demand for the good or service in question.

Basically, these six points accomplish two broad goals:

- ✔ They guarantee that people want to buy and sell in a market environment.
- ✔ They ensure that markets take into account all the costs and all the benefits of producing and then consuming a given amount of output.

We address each point separately in the next two sections.

Guaranteeing that people want to participate in markets

The requirement that both buyers and sellers have access to full and complete information guarantees that both are willing to negotiate without having to worry that the other person has secret information. (In Chapter 15, we explain how markets break down if one side or the other has more information.)

The requirement that property rights be set up in such a way that buyers *have* to pay sellers ensures that there are sellers willing to provide the product. As a counter example, consider trying to sell tickets to an outdoor fireworks display. Because everyone knows that they can see the display for free, nobody wants to pay for a ticket. But if sellers can't sell tickets, they have no incentive to put on a display. (In Chapter 15, we discuss situations like these and how society must deal with them given that markets can't.)

Capturing all costs and benefits

The requirements that supply curves capture all costs and demand curves capture all benefits ensure that a proper cost-benefit calculation can be made. For example, if a steel factory can pollute for free, there's no way the price of steel can reflect correctly the damage that the factory's pollution does to the environment. On the other hand, if the government continuously forces the factory to pay for clean-up costs, these costs are going to be reflected in the market price, thereby allowing society to weigh the costs and benefits properly of the company's output. (Chapter 14 deals with ways to help markets along if supply and demand curves don't reflect all costs and benefits.)

If the first four requirements for free markets are met, market forces can still reach a social optimum only if they are free of interference. The fifth requirement eliminates problems like monopolies, in which individual buyers or sellers are so powerful that they can manipulate the market pricing in their own favour. The sixth requirement stipulates that supply and demand must be allowed to determine freely the market price and market quantity unimpeded by government-imposed price ceilings or floors. (In Chapter 8, we explain the problems with price ceilings and floors and discuss how they hurt society.)



If all six requirements are met, an amazing thing happens. Supply and demand automatically achieve the social optimum without the government or socially conscious activists having to do anything. This insight was the basis of Adam Smith's metaphor of an invisible hand that seems to guide markets to do the right thing despite nobody being in charge – and despite the fact that each individual in the market may well be looking out only for his or her own interests.

So take this insight to heart by looking out for your own interests and reading the rest of this chapter carefully. You may just end up promoting the social optimum.

Analysing the efficiency of free markets

Economists use supply and demand curves to demonstrate that free markets produce socially optimal levels of output. But the simple insight behind this result is that a unit of output can be socially beneficial to produce and consume only if the benefits that people derive from consuming it exceed the costs of producing it.

This simple idea is, in fact, why demand curves and supply curves are so useful in analysing the social optimum. As we explain in Chapter 8, demand curves quantify the benefits that people get from consumption by showing what they're willing to pay to consume each and every particular unit of output. In a similar fashion, in Chapter 10 we explain how supply curves quantify the cost of producing each and every particular unit of output.

Using supply and demand to compare costs and benefits

By drawing the demand and supply curves for a good or service together on the same graph, you can easily compare the benefits and costs of producing each and every unit of output. To see how this is done, take a look at Figure 11-1, on which we draw a demand curve, D , and a supply curve, S .

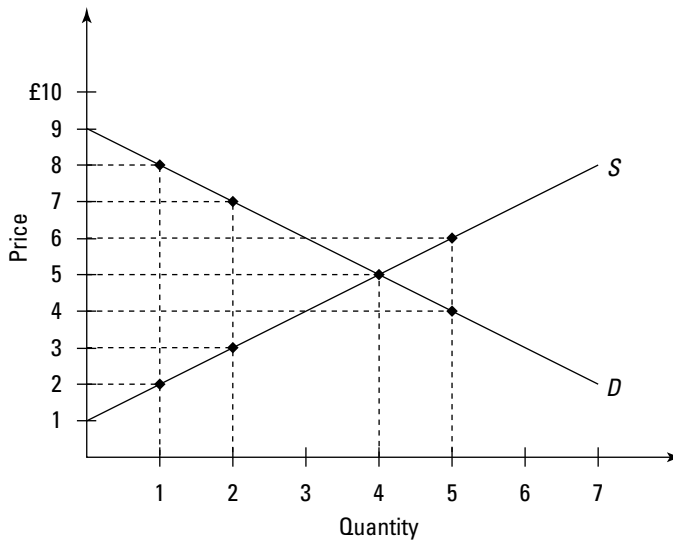


Figure 11-1:
Comparing
costs and
benefits
using supply
and demand
curves.

To start, take a look at one unit of output on the horizontal axis. At that output level, go vertically up to the demand curve and see that people are willing to pay £8 for one unit of output. At the same time, go vertically up to the supply curve to see that firms are willing to supply one unit at a cost of £2.

Putting these facts together, you can see that producing this first unit of output is socially beneficial because the value is worth more to buyers (£8) than it costs sellers to produce (£2). Put slightly differently, although the resources to make this unit of output cost society £2, those resources bring £8 in benefits when converted into this particular good or service. Because the benefits exceed the costs, this unit of output should be produced.

Now look at the second unit of output. Going vertically up to the demand curve tells us that people are willing to pay £7 for that unit, whereas going vertically up to the supply curve tells us that the second unit costs £3 to produce. Again, benefits exceed costs. Again, this unit of output should be produced.

In contrast, look at the fifth unit of output. By going up vertically, you can see that the costs as given by the supply curve for producing the fifth unit are £6, whereas the benefits as given by the demand curve are only £4. Because the costs of producing this unit exceed what anyone is willing to pay for it, this unit of output shouldn't be produced.

In other words, producing the fifth unit of output destroys value, because making it involves converting £6 worth of resources into something worth only £4 to consumers. Producing it destroys wealth.

Determining the socially optimal output level

The next thing to notice is that Figure 11-1 can tell you precisely what quantity (q) of output should be produced, because the supply and demand curves let you quickly compare costs and benefits for every possible output level.

Only three cost-benefit relationships exist:

- ✓ For every bit of output where $q < 4$, benefits exceed costs.
- ✓ At exactly $q = 4$ units, benefits equal costs.
- ✓ For every bit of output where $q > 4$, costs exceed benefits.

Economists look at these relationships and conclude that the socially optimal level of output to produce is $q = 4$ units, because for these units benefits exceed costs or are at least equal to costs. By producing the first four units of output, society gains or is at least not made any worse off.

The socially optimal output level is always devastatingly easy to identify on any supply and demand graph: just look to the quantity produced where the demand and supply curves cross.

Realising that free markets produce the socially optimal output level

Adam Smith's big insight was to realise that free markets produce exactly the socially optimal output level on their own without anyone having to direct them to do the right thing.

The proof of this fact is almost trivial. All you have to do is look at Figure 11-1 and realise that the market equilibrium quantity – which happens when the market price is free to adjust so that the quantity supplied by sellers equals the quantity demanded by buyers – is determined by where the supply and demand curves cross. (To understand why, see Chapter 8.) The market equilibrium quantity is four units of output, which is exactly how many units you want to produce if you are using the demand and supply curves to compare benefits and costs.

Now, here's the clever bit. Sensible business people are not in the, er, business of producing things when the cost of production is greater than the value received from sales. In this case, the calculation is simply that it costs £6 to produce and you'd make £4 from selling. Therefore you wouldn't do it. The key thing to understand is that it needn't take any central planner or civil servant to determine the right level of output. Firms that are free to adjust their output to market conditions will take the decision in response to conditions like these.

This amazing result greatly simplifies life, because it eliminates the need to have a government official or any other sort of central planner constantly checking to see if exactly the right amount of output is being produced.

Using total surplus to measure gains



Economists use a concept called *total surplus* to total up the gains that come from producing the socially optimal output level. The gain, or surplus, comes from the fact that benefits exceed costs for the units of output that are produced.

The total surplus turns out to be divided between consumers and producers. The part of the total surplus that goes to consumers is (naturally) called *consumer surplus*, whereas the part that goes to producers is called *producer surplus*.

In the sections that follow, we tackle consumer surplus first and then move on to producer surplus. After we explain each separately, we add them together to explain total surplus. (And we hope that when you're done with this section, you feel like you've received at least a *little* consumer surplus.)

Measuring the consumer surplus of a discrete good

Consumer surplus is the gain people receive when they can buy things for less than what they're willing to pay.

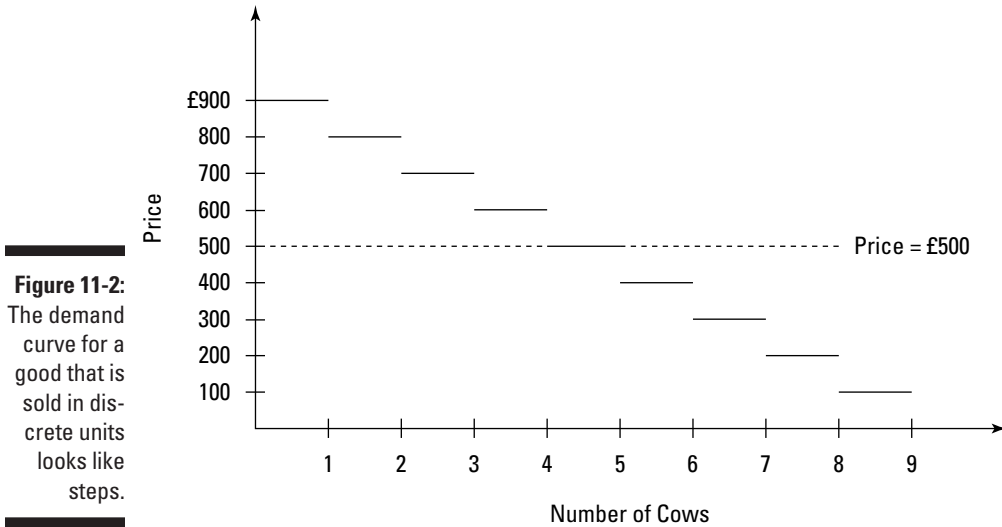
The easiest way to understand consumer surplus is to first look at a discrete good. A *discrete good* is a good that comes only in discrete units. For example, you can buy one car or 57 cars, but you can't buy 2.33 cars. You can purchase one horse or thirteen cows but not fractional amounts of livestock (at least if you want them alive!).



Look at Figure 11-2, which shows the demand for cows. Because cows come in discrete units, you don't get a smooth, downward-sloping curve. Instead, you get what mathematicians call a *step function*. This means, for example, that people are willing to pay £900 for the first cow, £800 for the second cow, £700 for the third cow and so on.

Now imagine that the market price of cows is £500, which is why we draw a horizontal dotted line at that price. Compare that price with what people are willing to pay for each cow.

For the first cow, people are willing to pay £900. Because the market price of cows is only £500, these buyers come out ahead because they're able to purchase a cow for £400 less than they're willing to pay. Or, as economists like to say, the *consumer surplus* on the first cow is £400.



Next, look at the second cow. People are willing to pay £800 for it, but because the market price is only £500, they receive a consumer surplus for that cow of £300.

Similarly, for the third cow, people get a consumer surplus of £200, because they're willing to pay £700 for it but have to pay only the market price of £500.

For the first four cows, a positive consumer surplus exists, whereas on the fifth cow people just break even because they're willing to pay £500 and the cow costs £500. This means that people want to buy only five cows. (Economists always assume that when the price equals your willingness to pay, you go ahead and buy.)

To calculate consumer surplus for a discrete good such as cows, we need to total the surpluses that people get on each unit that they choose to buy. In this case, the total is £1,000 (£400 for the first cow, plus £300 for the second cow, plus £200 for the third cow, plus £100 for the fourth cow, plus £0 for the fifth cow).

To show this £1,000 of consumer surplus in the graph in Figure 11-3, we shade in the area below each step and above the horizontal price line at £500. The staircase-shaped area equals £1,000.

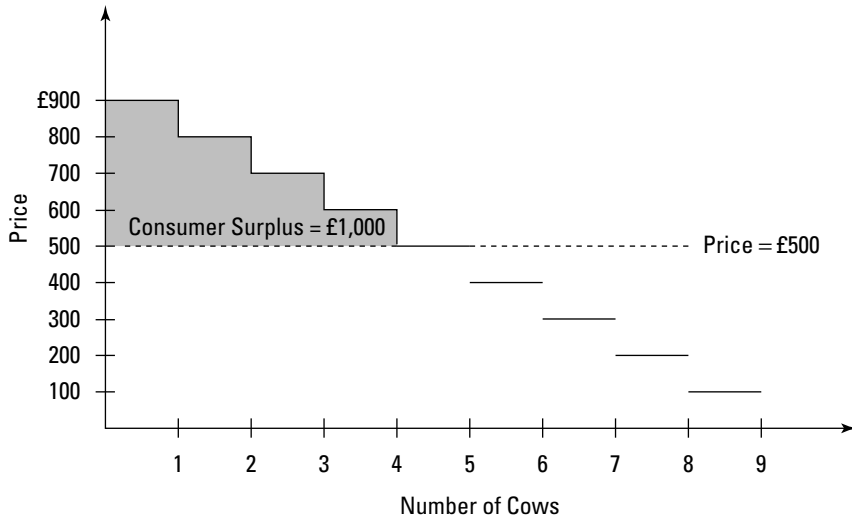


Figure 11-3:
Consumer surplus for a discrete good.

Measuring the consumer surplus of a continuous good

Consumer surplus can also be computed for continuously measured goods and services – things like land, cooking oil or hours of music lessons, which aren't necessarily sold in discrete units. In other words, you can buy fractional amounts of continuously measured goods, such as 78.5 acres of land, 6.33 litres of cooking oil or 2.5 hours of music lessons.

The demand curves for continuously measured goods are much nicer than the step functions that you get for discretely measured goods. In fact, the demand curves for continuously measured goods are the smooth, downward-sloping lines that you're used to seeing (such as in Chapter 8).

The smoothness of such demand curves means that when you graph consumer surplus for a continuously measured good, you get a triangular area that lies below the demand curve and above the market price. You can see this wedge illustrated in Figure 11-4, which depicts the cooking oil market.



In Figure 11-4, the price of cooking oil is £5 per litre. At that price, people want to buy 1,000 litres of cooking oil. The demand curve lies above the horizontal £5 price line, which means that buyers are made better off by buying these 1,000 litres because they are worth more to the buyers than the £5 per litre that it costs to buy them.

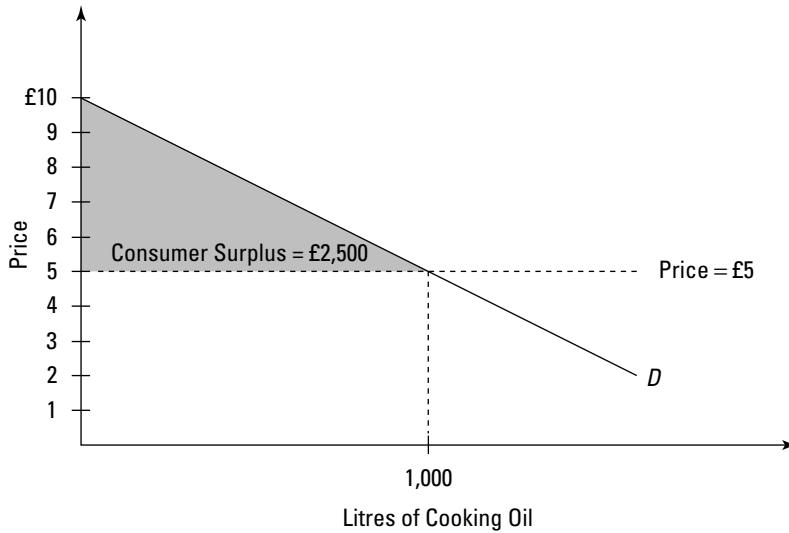


Figure 11-4:
Consumer surplus for a continuous good.

To calculate consumer surplus for a continuous good, you total up all the gains that people receive when buying for less money than they are willing to pay – just as for a discrete good. But because we’re now dealing with a triangle, totalling up requires a bit of geometry. Don’t worry, the method’s easy. You simply use the formula for the area of a triangle ($\frac{1}{2} \times \text{base} \times \text{height}$) to find the total surplus. In this case, you multiply $\frac{1}{2} \times 1,000 \times 5 = \text{£}2,500$.

Measuring producer surplus

Producer surplus measures the gain that firms receive when they can sell their output for more than the minimum price that they were willing to accept. You can calculate producer surplus for both discrete and continuous goods, just as you can calculate consumer surplus for each. In this section, we offer an example of calculating producer surplus for a continuous good.

To get a good handle on producer surplus, take a look at Figure 11-5, which shows the supply curve, *S*, for cooking oil. This supply curve is crucial for determining producer surplus because each point on the supply curve tells you the minimum that you have to pay suppliers for them to supply you with the associated amount of output. By comparing each minimum value with the higher market price that they actually receive when they sell their output, you can compute producer surplus. (For more on supply curves and how to interpret them, see Chapter 8.)

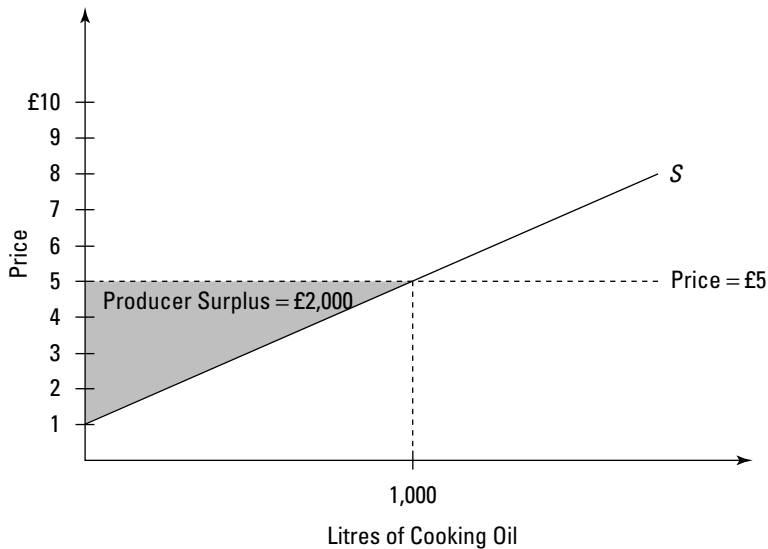


Figure 11-5:
Producer surplus for a continuous good.

The price of cooking oil is still £5 per litre. And the way we draw the graph, producers are going to want to supply exactly 1,000 litres of cooking oil at that price. They want to supply this much because for each drop of oil up to and including the very last drop of the 1,000th litre, the production costs as given by the supply curve are less than the £5 per litre that producers get when they sell the oil.

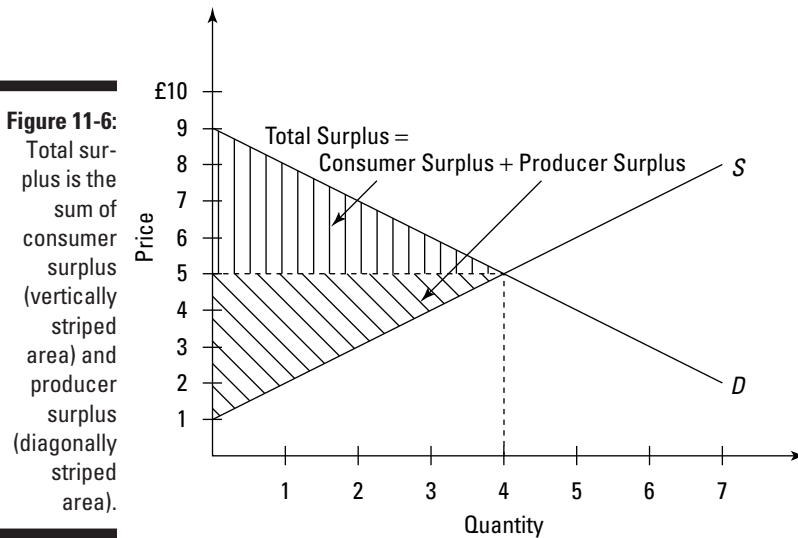
But, crucially, producers are willing to supply almost all that cooking oil for *less than* the £5 per litre market price. You can see this by the fact that the supply curve lies below the horizontal price line up to the very last drop of the 1,000th litre. The fact that producers receive £5 per litre for all the oil, despite being willing to produce it for less, is the source of the producer surplus, which is represented by the area of the shaded triangle.

Using the formula for the area of a triangle ($\frac{1}{2} \times \text{base} \times \text{height}$), you can compute that the producer surplus in this example is £2,000. Producers are £2,000 better off by selling the 1,000 litres of oil because the total cash they get from selling these 1,000 litres exceeds the minimum amount that they were willing to accept by £2,000.

Computing total surplus

The *total surplus* that society receives from producing the socially optimal level of output of a certain good or service is simply the sum of the consumer surplus and producer surplus generated by that output level.

Figure 11-6 illustrates total surplus for a market in which the equilibrium price and quantity are, respectively, $p^* = £5$ and $q^* = 4$. (If this graph looks familiar, that's because it resembles Figure 11-1.)



We draw the total surplus area so that you can clearly see that it consists of consumer surplus plus producer surplus. The horizontal line extending from the market equilibrium price (£5) separates the two parts. The consumer surplus triangle is filled with vertical lines, whereas the producer surplus triangle is filled with diagonal lines.

Again using the formula for the area of a triangle, we multiply $\frac{1}{2} \times 4 \times 8$ to figure out that for this graph the total surplus is £16. The total gain to society of producing at this output level is £16.

Contemplating total surplus

Total surplus is very important because it puts a number on the gains that come from production and trade. Firms make things to make a profit. People spend money on things because consuming those things makes them happy. And total surplus tells you just how much better off both consumers and producers are after interacting with each other.

By putting a number on the gains made by their interaction, total surplus also provides a benchmark by which economists can measure the harm that comes from government policies that interfere with the market. For example, saying that price subsidies hurt consumers is one thing, but being able to say by exactly how many pounds consumers are harmed is quite another. And that's the subject we cover next.

When Free Markets Lose Their Freedom: Dealing with Deadweight Losses

As we note earlier in the chapter, economists love free markets because free markets produce only those units for which benefits exceed costs. In other words, the market equilibrium ensures that total surplus is as large as possible.



Anything that interferes with the market's ability to reach the market equilibrium and produce the market quantity reduces total surplus. Economists use the colourful term *deadweight loss* to refer to the amount by which total surplus is reduced.

In the sections that follow, we give you detailed examples of deadweight losses caused by price ceilings and taxes. These types of market interference are both under the government's control, but don't think that *only* government policy causes deadweight losses. Anything that reduces output below the market quantity causes a deadweight loss. Monopolies and oligopolies can be to blame, as can asymmetric information, poorly allocated property rights and public goods problems – all things that we discuss in the next few chapters.

Dissecting the deadweight loss from a price ceiling

As an example of a deadweight loss, look at Figure 11-7 in which the government has imposed a price ceiling at p^c . As we discuss in Chapter 8, *price ceilings* are maximum prices at which sellers can legally sell their product. Generally, price ceilings are intended to help buyers obtain a low price, but, as we're about to show you, they can cause a lot of harm.

To see the damage that price ceilings can inflict, first notice that at a maximum price of p^c , suppliers are going to want to sell only q^l units of output (the *L* stands for *low*). In other words, at that price, only the first q^l units of output are profitable to produce. By contrast, if no price ceiling exists and the market is left to its own devices, suppliers choose to produce the market equilibrium quantity of output, q^* .

Consequently, if this were a free market, the total surplus would be represented graphically by the triangle defined by points *A*, *B* and *C*. But because only q^l units of output can be produced, the total surplus area is reduced down to the shaded area with corners at *A*, *B*, *F* and *E*.

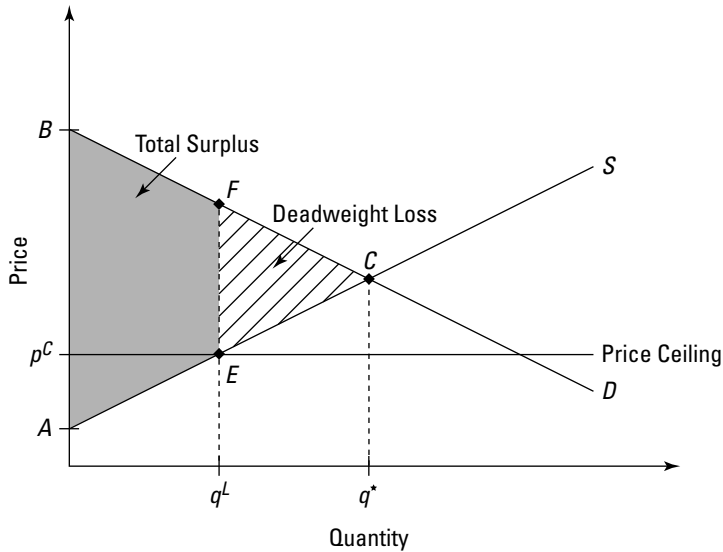


Figure 11-7:
Reduced total surplus caused by a price ceiling.

The difference between the total surplus generated by producing q^* versus q^L units of output is the diagonally striped triangle defined by points E , F and C . The area of this triangle illustrates the deadweight loss that comes from reducing output below the socially optimal level, q^* .

The price ceiling is harmful because for all units between q^L and q^* benefits exceed costs, meaning that such units should be produced. By tallying up the gains that would come from producing and consuming these units, the deadweight loss triangle can precisely measure the harm that results from interfering with the market. If it doesn't appear as if that's the case, consider the supply of rented homes. A maximum rent will – *other things being equal* – reduce the supply of homes that landlords are willing to rent out. The deadweight loss incorporates that loss of supply.

Analysing the deadweight loss of a tax

Taxes on goods and services also cause deadweight losses. This situation happens because such taxes raise the costs of producing and consuming output. When these costs are artificially raised by a tax, people respond by producing and consuming fewer units of output than they did before the tax was imposed. Because each unit consumed before the tax was imposed was a unit for which benefits exceeded costs, the reduction in output that results from the tax necessarily reduces total surplus and causes a deadweight loss.



Seeing how a tax shifts the supply curve

Before we discuss in more detail the deadweight loss that results from a tax, we have to show you that imposing a tax on the seller shifts supply curves vertically by the amount of the tax. Consider this example – the supply of beef in a beef market in which the government is going to impose a tax of £1 per kilo.

Figure 11-8 shows two curves. (Well, actually they're straight lines, but play along with us for a moment here.) The lower one, S , is the supply curve for beef. The higher one, labelled $S + tax$, is the supply curve after the tax is imposed. The important thing to realise is that the curve $S + tax$ is simply the original supply curve shifted up vertically by the amount of the tax, which in this case is £1.

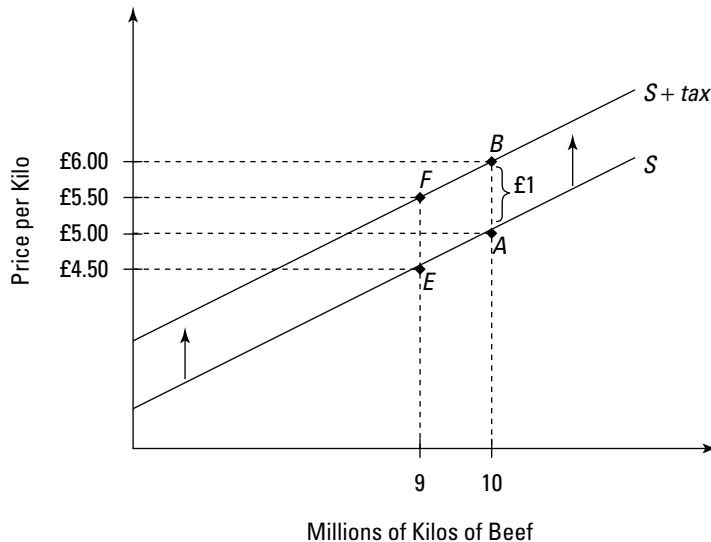


Figure 11-8: Imposing a £1 tax on beef shifts the supply curve vertically by £1, from S to $S + tax$.

The reason why the supply curve shifts up vertically by the amount of the tax is connected with the motivation of suppliers. In Chapter 8, we explain that each point on the supply curve tells you the minimum that you would have to pay suppliers to get them to supply the relevant quantity. For example, look at point A . Because point A is on the supply curve, you know that you have to pay £5 per kilo if you want suppliers to provide 10 million kilos of beef. Similarly, point E tells you that you have to pay suppliers £4.50 per kilo if you want them to supply only 9 million kilos of beef.

If the government comes in and imposes a tax of £1 per kilo, it affects how much you have to pay the suppliers to motivate them. If you still want 10 million kilos of beef, you have to pay the original amount required to motivate the suppliers to supply you with that much beef (£5 per kilo), as well as enough money to pay the taxes on that much beef (£1 per kilo).

Graphically, this means that point *A* on supply curve *S* shifts up by the £1 amount of the tax to become point *B* on the *S + tax* curve. For the same motivational reasons, point *E* on the supply curve must shift up to point *F* on the *S + tax* curve. That is, if you have to pay suppliers £4.50 per kilo to motivate them to supply you with 9 million kilos of beef in a world in which £1 per kilo must go to the government in taxes, you have to collect a total of £5.50 per kilo. And that's exactly what happens at point *F*.

Every point on the supply curve, *S*, must shift up vertically in the same way that points *A* and *E* do, so the *S + tax* curve captures what the supply curve looks like after the tax is imposed. With this shift in mind, you're ready to discover how this sort of taxation causes deadweight losses.

Seeing how a tax causes deadweight losses

Figure 11-9 adds a demand curve, *D*, to Figure 11-8 so that you can see what happens to total surplus when the government imposes a £1 per kilo tax on the beef that's sold in the beef market.

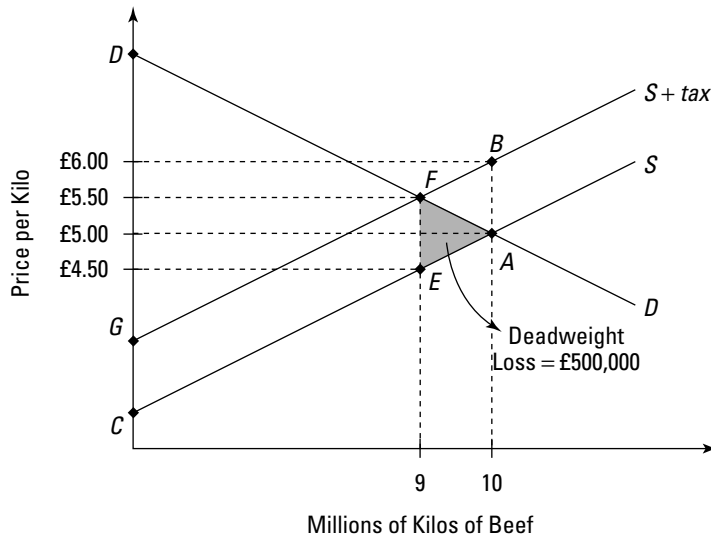


Figure 11-9:
The deadweight loss caused by a £1 tax on beef.

Before the tax, the market equilibrium happens at point *A*, where supply curve *S* crosses demand curve *D*. At that point, producers supply 10 million kilos of beef at a price of £5 per kilo. The total surplus in this case is given by the triangle defined by points *C*, *D* and *A*.

After the tax is imposed, however, the equilibrium happens at point *F*, where the *S* + *tax* curve crosses the demand curve. At that point, the price of beef is £5.50 per kilo, and 9 million kilos are supplied. (Only 9 million kilos are supplied because after the government takes its £1 per kilo in taxes, only £4.50 remains to motivate suppliers. You can see from the supply curve that at that much money per kilo, suppliers want to supply only 9 million kilos.)

Because of the tax, the amount of beef supplied falls from 10 million kilos to 9 million kilos. Furthermore, the total surplus is reduced to the triangle whose three corners are *G*, *D* and *F*.

You can immediately see that this new total surplus is much smaller than the old total surplus. But before we start ranting about the evils of government, we need to take account of the fact that taxes are being collected. Taxes (theoretically, at least) benefit society, so we need to include this amount when calculating the total surplus of this good sold at this price. At the new equilibrium, £9 million in taxes are collected because the 9 million kilos of beef sold are taxed at £1 each.

The £9 million in tax collections are represented graphically by the parallelogram whose corners are *C*, *G*, *F* and *E*. This area was previously contained in the old total surplus triangle whose corners were *C*, *D* and *A*. Consequently, the area that used to be part of the old total surplus hasn't been destroyed, but merely transferred to the government.

However, part of the old total surplus *has* been destroyed. This part is shown graphically by the shaded deadweight loss triangle (with corners at *E*, *F* and *A*). This area captures the fact that society is made worse off by the reduction in beef output from 10 million kilos to 9 million kilos. (Okay, *now* you can start ranting about the evils of government.)

Measuring the size of the deadweight loss using the formula for the area of a triangle ($\frac{1}{2} \times \text{base} \times \text{height}$) tells us that the tax leads to a deadweight loss of £500,000. That's a big number representing a huge reduction in total surplus deriving from the fact that for each of the 1 million kilos of beef no longer being produced, benefits exceeded costs. All those gains are lost when the tax is imposed.



Deadweight losses are called deadweight losses because you can't say 'Your loss is my gain' in this situation. Benefit hasn't passed from consumers to producers, but instead the total level of benefit to society as a whole (in this case) is lower. We aren't talking about something that passes from one person to another. Deadweight losses are losses in the sense of annihilation. The gains that would have resulted if those million kilos of beef had been produced simply vanish; they are a dead weight that we must bear in our efforts to maximise human happiness given our limited resources.

Hallmarks of Perfect Competition: Zero Profits and Lowest Possible Costs

Earlier in this chapter, we demonstrate that free markets produce only units of output for which benefits are at least as great as costs. Another wonderful thing about free markets and competition is that output is produced at the lowest possible cost.

This fact is extremely important because it means that free markets are as efficient as possible at converting resources into the goods and services that people want to buy.

In addition, markets save society a lot of money because they produce efficiently without requiring any human intervention. We don't have to pay big salaries to experts to make sure that markets run efficiently; markets do the job for free.

Understanding the causes and consequences of perfect competition

To ensure that markets function efficiently, you need really strong competition between firms, a situation that economists refer to as *perfect competition*.

As we explain in Chapter 10, perfect competition exists when many firms within a given industry are all producing identical (or nearly identical) products. The following things are also true when perfect competition exists:

- ✓ Every firm is a *price taker* – that is, has to accept the market equilibrium price for what it produces – because its output is a very small fraction of the industry's total output (see Chapter 10).
- ✓ Every firm has identical production technology.
- ✓ Every firm is free to enter or leave the industry as it pleases.

When these requirements are met, perfect competition leads to two excellent outcomes:

- ✓ Every firm in the industry makes zero economic profits.
- ✓ Every firm produces output at the minimum possible cost.

The first outcome doesn't mean that businesses earn no money above the costs of doing business; otherwise, no one would go into business. Firms must earn enough money to keep entrepreneurs motivated to stay in business (and to attract other entrepreneurs to open new firms).

So what does the first outcome mean? In Chapter 10, we explain that the *economic profits* earned by a firm are any monies collected above and beyond what is required to keep an entrepreneur owner interested in continuing in business. So the fact that perfect competition leads to zero economic profits means that firms just barely want to stay in their industry.

It also means that nobody in the industry is getting filthy rich at anyone else's expense. Rather, they're doing just well enough to keep on supplying the output that society wants them to supply. This situation is great for society, because paying entrepreneurs more than necessary to get them to do what society wants is wasteful.

We discuss the second outcome of perfect competition – the fact that firms all end up producing output at the lowest cost possible – in the upcoming section 'Graphing how profits guide firm entry and exit'. This outcome is also good for society because it means that the least possible amount of resources are consumed while making the output that society wants produced.

Peering into the process of perfect competition

The previous section gives you an idea of how perfect competitive markets emerge and how they benefit society. But how does perfect competition actually work? The following four steps explain:

1. The market price of the output sold by every firm in the industry is determined by the interaction of the industry's overall supply and demand curves.
2. Each of the firms takes the market price as given and produces whatever quantity of output maximises its own profit (or minimises its own loss if the price is so low that making a profit isn't possible).

3. Because each firm has an identical production technology, each chooses to produce the same quantity and consequently makes the same profit or loss as every other firm in the industry.
4. Depending on whether firms in the industry are making profits or losses, firms enter or leave the industry until the market price adjusts to the level where all remaining firms are making zero economic profit.

The fourth point in this process – firm entry and exit – is very important. To understand this point clearly, break it into two cases, one where every firm in the industry is making a profit because the market price is high, and another where every firm in the industry is making a loss because the market price is low:

✓ **Attracting new firms by making profits:** If every firm in an industry is making a profit, new firms are attracted to enter the industry too, in hopes of sharing the profits. But when they enter, total industry output increases so much that the market price begins to fall. As the price falls, profits fall, thereby lowering the incentive for more firms to enter the industry.

The process of new firms entering the industry continues until the market price falls so low that profits drop to zero. When that happens, the incentive to enter the industry disappears, and no more firms enter.

✓ **Losing existing firms when making losses:** If every firm in an industry starts out making losses because the market price is low, existing firms that can't stand losing money exit the industry. When they do, total industry output falls. That reduction in total supply, in turn, causes the market price to rise. And as the market price rises, firms' losses decrease.

The process of firms leaving and prices rising continues until the remaining firms are no longer losing money.

As we explain in the preceding section, the fact that firms can freely enter or leave the industry means that after all adjustments are made, firms always make a zero economic profit. In other words, if perfect competition exists, you don't have to worry about firms exploiting anyone; they just barely make enough money to stay in business.

The other important result of perfect competition – that competitive firms produce at minimum cost – becomes apparent if we flesh out the four-step process of perfect competition by using the cost curves that we explain in Chapter 10. If you haven't read that chapter, this section may cause your eyes to cross (and you know what your mother said about the dangers of crossing your eyes). We encourage you to take a look at Chapter 10 before moving on to the next section.

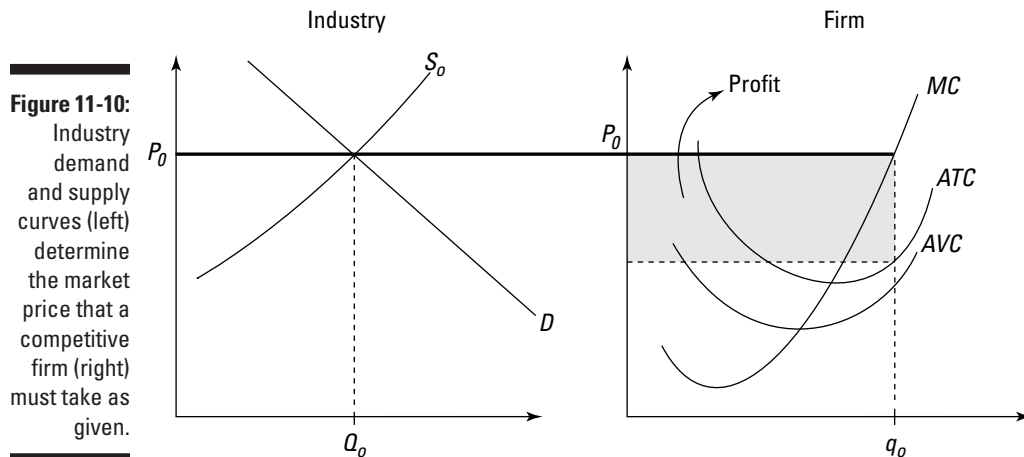
Graphing how profits guide firm entry and exit

In this section, we use the firm cost curves that we introduce in Chapter 10 to demonstrate how market forces automatically cause firms to produce output at the lowest possible cost. To make this process clear, we present two cases. In the first, firms begin by making profits. In the second, firms begin by making losses. Either way, adjustments happen so that they end up making zero economic profits and producing at minimum costs.

Visualising firm entry when profits exist



To see how an industry adjusts when it starts off making profits, look at Figure 11-10, which consists of two graphs. The one on the left gives the market demand curve, D , and the initial market supply curve, S_0 , for tennis balls. The one on the right gives the cost curves for one of the many identical firms that make tennis balls.



Because the firms in this industry are identical, they all have the same cost structures. In particular, they all have the same marginal cost curve (MC). This point is important because, as we show in Chapter 10, a competitive firm's marginal cost curve is its supply curve.

The firm in our example takes the market price, P_0 , determined by supply and demand in the left graph, and uses it to work out its profit-maximising output level in the right graph. (To emphasise that P_0 is the same in both graphs, we draw a solid horizontal line that goes all the way across both graphs.)

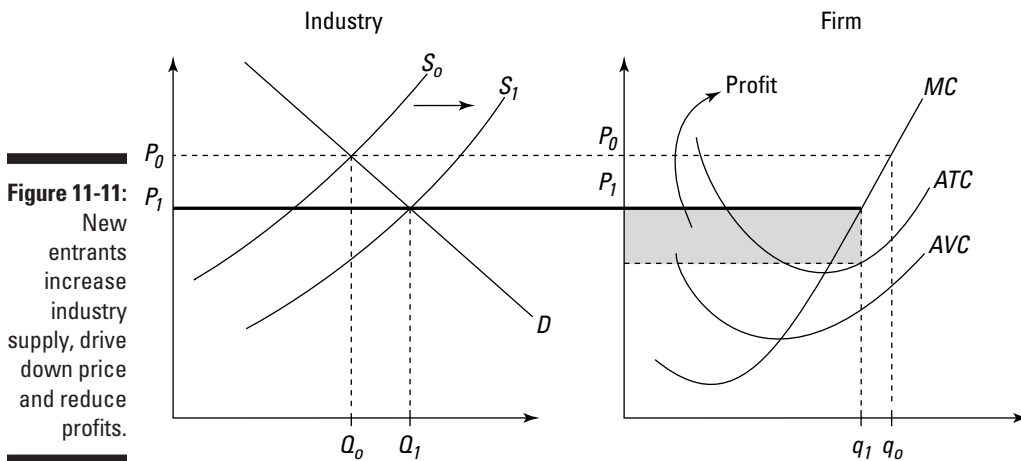
As we show in Chapter 10, each firm chooses to produce the output level at which the horizontal price line intersects the MC curve. In the right-hand graph, we label the output level q_o . In the left-hand graph, you can see that the industry's total supply is Q_o . The industry's total supply is simply each individual firm's output, q_o , times the total number of firms in the industry.

Next, focus on the fact that each firm runs a profit when the market price is P_o . The profit is shown by the shaded rectangle in the right graph. (We explain these profit rectangles in – where else? – Chapter 10.)

This profit is important because it attracts entrepreneurs to enter the industry. They realise that they can set up yet more identical firms and make some nice profits. As economists like to say, profits attract entrants.

Seeing how new entry reduces profits

Figure 11-11 shows what happens when the new entrants to the industry arrive. Their new production increases overall production so that the total supply curve shifts from S_o to S_1 in the left-hand graph, lowering the market equilibrium price from P_o to P_1 .



Each of the price-taking firms reacts to the lower price by producing a lower output level, q_1 , which you can see illustrated in the right-hand graph. More importantly, the firms' profits decrease, which you can see by comparing the shaded profit rectangles in Figures 11-10 and 11-11.

The new entry results in smaller profits. The smaller profits are less attractive to entrepreneurs. So although new entry continues – because some profits are still available – that new entry is less than when profits were larger.

Seeing how enough entry drives profits to zero

What ends up happening, in fact, is that entry continues until prices fall so far that all profits are driven away. This situation is illustrated in Figure 11-12, in which new entry increases supply still more, to S_2 . The result is that the market price falls to P_2 , which results in zero profits. (Note that no shaded profit rectangle exists in the right-hand graph.) Because profits fall to zero, entry ceases.

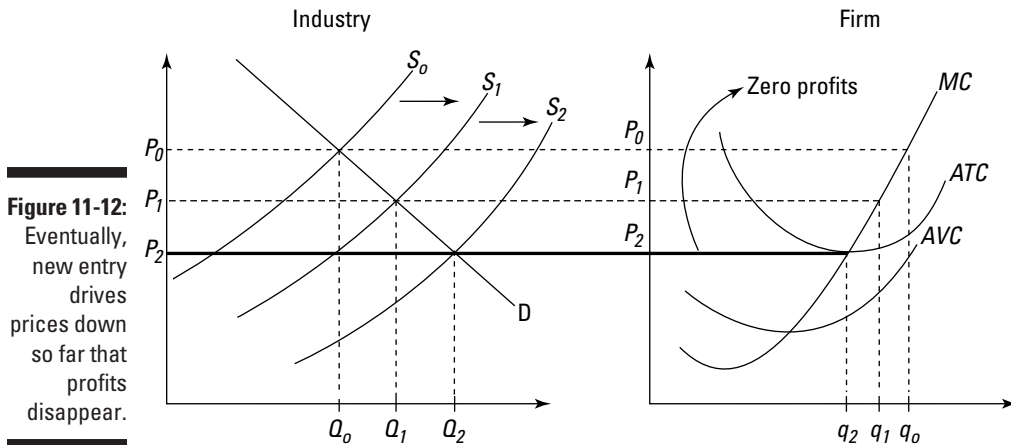


Figure 11-12: Eventually, new entry drives prices down so far that profits disappear.

Realising that zero profits also mean minimum cost production

When profits are driven to zero by the entry of new firms, the cost per unit at which output is produced is minimised. You can see this fact in the right-hand graph of Figure 11-12: notice that when faced with price P_2 , firms choose to produce at the quantity that minimises per-unit production costs.

The output that firms choose to produce, q_2 , lies exactly at the minimum point of the U-shaped average total cost curve (ATC). When output is produced at that level, the average cost per unit is lower than at any other output level. (In other words, any other output level results in a higher average total cost.) This situation is wonderful because it means that each firm is being as efficient as possible, producing output at the lowest possible cost per unit. Moreover, each firm is voluntarily choosing to produce at that level without any need for coercion.



Profits serve as a self-correcting feedback mechanism. High profits automatically attract new entrants who automatically increase supply and drive prices down. That process continues until no more profits and no more new entrants exist. But more importantly, it continues until each and every firm is producing output at the most efficient, least-cost output level. This is truly Adam Smith's invisible hand at work.

Visualising firm exit when losses exist

A similar feedback mechanism leads to zero profits and efficient production if the industry starts out making losses. To see this, take a look at Figure 11-13, where the initial supply curve, S_3 , interacts with the demand curve, D , to produce a very low market price of P_3 .

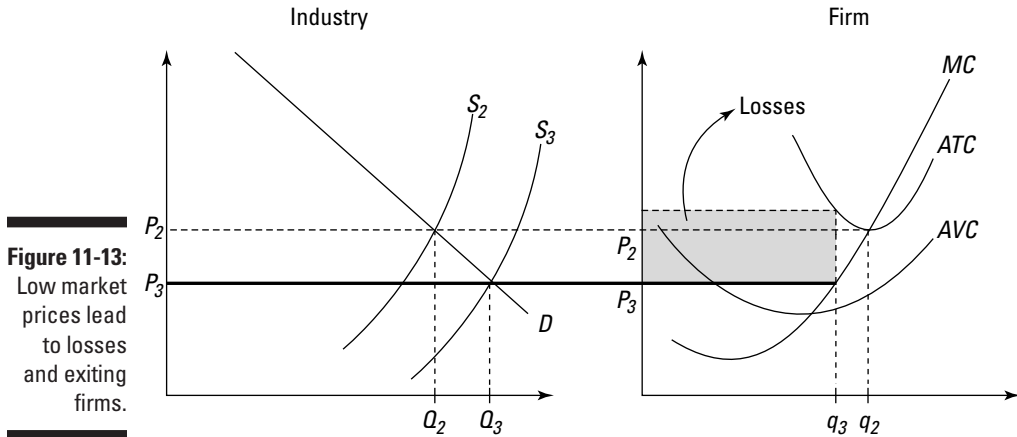


Figure 11-13: Low market prices lead to losses and exiting firms.

At this market price, you can see in the right-hand graph that each firm in the industry is making a loss, which is shown by the shaded rectangle.

This loss discourages all the firms in the industry, and those in the weakest financial condition begin to exit. As that happens, the industry supply curve in the left-hand graph shifts left (because supply decreases). That shift raises the market price and reduces the losses made by firms remaining in the industry. But as long as losses exist, firms continue to exit until the supply curve moves all the way back to S_2 , at which point the market price is P_2 , and firms are making zero profits as in Figure 11-12.

When the market price reaches P_2 and firms are making zero profits, the exit of firms stops and, more importantly, each firm is producing at the least-cost output level, q_2 .

Understanding that entry and exit don't happen instantly

You've now seen that market pressures always push perfectly competitive firms to produce at the lowest possible per-unit cost. Keep in mind that this nice result doesn't happen overnight. Whether firms are making profits or sustaining losses, it takes time for new firms to enter (if profits are available) or for existing firms to leave (if losses exist).



Depending on the industry, these adjustment processes may take anywhere from a few weeks to a few years. For example, setting up new power plants takes some time because building a new power plant takes at least a year. Similarly, even if agricultural prices fall and farmers are making losses, those farmers who drop out of the industry don't do so until the next growing season. On the other hand, if producing St George's cross flags suddenly becomes really popular (such as when England is playing at a major football tournament), you can be sure that scores of new firms pour into the industry within weeks.

The wonderful thing about perfect competition is that market forces are always acting to drive firms to produce at the minimum possible cost. As we show you in the next few chapters, this lovely result falls apart when monopolies, oligopolies, public goods and other problems prevent or preclude perfect competition.

Chapter 12

Monopolies: How Badly Would You Behave If You Had No Competition?

In This Chapter

- ▶ Producing less and charging more than competitive firms
 - ▶ Maximising profit
 - ▶ Benefiting society (in certain situations)
 - ▶ Abiding by regulations
-

A firm that has no competitors in its industry is called a *monopoly*. Monopolies are much maligned because their profit incentive leads them to raise prices and lower output in order to squeeze more money out of consumers. As a result, governments typically go out of their way to break up monopolies and replace them with competitive industries that generate lower prices and higher output.

At the same time, however, governments also very intentionally create monopolies in other situations. For example, governments issue patents, which give monopoly rights to inventors to sell and market their inventions. Similarly, the BBC was originally a monopoly created by government (although subsequently opened up to some competition).

In this chapter, we explain why society forbids monopolies in some situations and promotes them in others. First, we show you that profit-maximising monopolies compare unfavourably with competitive firms, because they set higher prices and produce less output than competitive firms. Then, we explain how these problems may, in certain cases, be outweighed by other factors – the need to promote innovation, for example, and the odd fact that in some cases having a lot of competitors is just too annoying.

All interesting stuff, so we expect to have a monopoly on your attention! (You have our permission to groan now.)

Examining Profit-Maximising Monopolies

Essentially, this chapter is one big exercise in cost-benefit analysis. Monopolies aren't all evil. Neither are they utterly good. Whether you want to have one in any particular instance depends on whether, in that situation, the benefits outweigh the costs.

This section goes into detail about the costs associated with monopolies. When we finish our cost analysis, we move on to the benefits of monopolies. This chapter helps you understand why society ruthlessly forbids monopolies in some industries while enthusiastically endorsing them in others.

Zeroing in on the problems that monopolies cause

In an industry with only one monopoly firm rather than lots of small competitive firms, three socially harmful things occur:

- ✓ The monopoly firm produces less output than firms in a competitive industry.
- ✓ The monopoly firm sells its output at a higher price than if the industry was competitive.
- ✓ The monopoly firm's output is produced less efficiently and at a higher cost than the output produced by firms in a competitive industry.

Although all these things are harmful to consumers, keep in mind that monopolies don't do these things for the sake of it. Instead, these outcomes are simply the result of monopolies acting to maximise their profits – which is, of course, the very same thing that competitive firms try to do.



Consequently, the difference in outcomes between a competitive industry and a monopoly industry doesn't have anything to do with bad intentions: it results from the fact that monopolies are free from the pressures that lead competitive industries to produce the socially optimal output level (see Chapter 11). Without these pressures, monopoly firms can increase prices and restrict output to increase their profits – things that competitive firms would also love to do but can't.

The lack of competitive pressure also means that monopoly firms can get away with costly, inefficient production. You need to take this problem seriously when considering whether the benefits of a monopoly outweigh its costs. We talk more about this issue later in the chapter.

Identifying the source of the problem: Decreasing marginal revenues



All the bad outcomes generated by a monopoly derive from the same source: unlike a competitive firm that faces a horizontal marginal revenue curve, the monopoly faces a downward-sloping marginal revenue curve. (*Marginal revenue* is the increase in total revenue that comes from selling each successive unit of a product; see Chapter 10.) This simple fact causes monopolies to charge more, produce less and produce at higher costs than competitive firms.

How can one little curve cause such mayhem? A downward-sloping marginal revenue curve implies that each additional unit that the monopoly sells brings less revenue than the previous unit. For example, whereas the 10th unit sold may bring in £8, the 11th brings in only £3. Obviously, such a situation reduces the incentive to produce a lot of output.

This situation also stands in stark contrast to the marginal revenue situation facing competitive firms. As we explain in Chapter 11, competitive firms face horizontal marginal revenue curves, meaning that whether they sell 11 units or 11,000, each unit brings in the same amount of money. Naturally, that's much more of an inducement to produce a lot of output.

Facing down demand

Why is there such a difference between the marginal revenue curves facing monopolies and competitive firms? A monopoly is free to choose the price it wants to charge along the demand curve it faces for its product. A competitive firm, on the other hand, has to take the market price as given (as we explain in Chapter 11).

A monopoly firm can choose its price because, being the only firm in its industry, it controls all the output in that industry. As a result, it can create a relatively high price by producing only a few units, or it can induce a relatively low price by flooding the market. In contrast, each firm in a competitive industry is such a small part of its industry that its choice of output makes too small a difference in total output to cause price changes. (See Chapter 11 for more on why competitive firms can't affect prices.)

The monopoly's ability to control the price by altering its output level means that it has to step back and consider what output level to produce. Obviously, because its goal is profit maximisation, it has to figure out what level of output is going to maximise its profits.

It turns out that a monopoly's profit-maximising output level is defined by the same condition as that of a competitive firm: produce at the output level where the marginal revenue curve crosses the marginal cost curve.

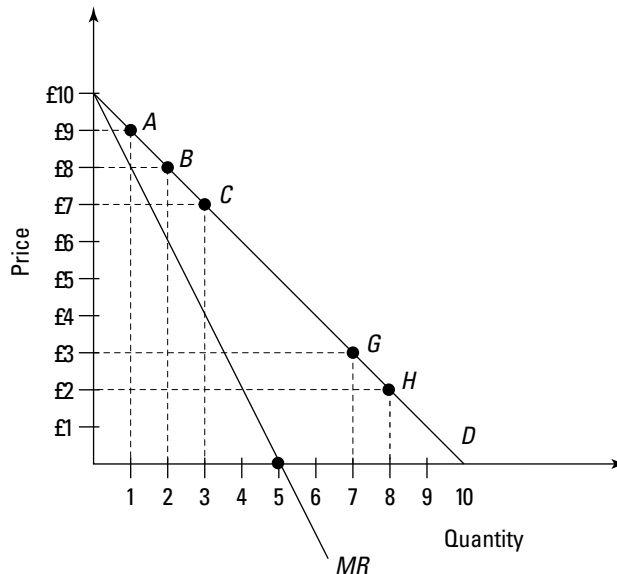
So the first step in working out how much a monopoly's going to produce is to figure out what its marginal revenue curve looks like. When you do, you can see where that curve crosses the monopoly's marginal cost curve to work out how much it's going to produce.

Deriving marginal revenue from the demand curve

A monopoly's marginal revenue curve has a precise relationship with the demand curve for the monopoly's output. The marginal revenue of each successive unit of output is less than the marginal revenue of the previous unit of output because demand curves slope downward. If the demand curve is a straight line, the slope of the marginal revenue curve is twice as steep as the slope of the demand curve, meaning that marginal revenue falls quite quickly as output increases.

To see how this works, take a look at Figure 12-1, in which we draw a demand curve and its associated marginal revenue (*MR*) curve.

Figure 12-1:
The *MR* curve for a monopoly facing a straight-line demand curve has a slope twice as steep as that of the demand curve.



We provide the data needed to draw these two curves in Table 12-1. The first column contains different output levels ranging from zero to ten units. The second column shows the price per unit that can be charged at each output level. The third column shows the total revenue that the monopoly would get for producing and selling each output level – the price per unit times the number of units. And the final column gives the marginal revenue – the change in total revenue – that happens as you increase output by one unit.

<i>Output</i>	<i>Selling Price</i>	<i>Total Revenue</i>	<i>Marginal Revenue</i>
0	10	0	
			9
1	9	9	
			7
2	8	16	
			5
3	7	21	
			3
4	6	24	
			1
5	5	25	
			-1
6	4	24	
			-3
7	3	21	
			-5
8	2	16	
			-7
9	1	9	
			-9
10	0	0	

To make it clear that marginal revenue represents the change in total revenue, the entries in the marginal revenue column are displayed between the two total revenue figures to which they correspond. For example, total revenue increases from £0 to £9 as you move from producing no output to one unit of output. That's why we place the marginal revenue of £9 at the top of the marginal revenue column, between the total revenue entries of £0 and £9.

As you can see in Figure 12-1, the marginal revenue (*MR*) curve starts at the same point as the demand curve, but it falls with twice the slope. It hits the horizontal axis at an output level of $q = 5$ rather than the $q = 10$ output level at which demand hits the horizontal axis (where q stands for quantity produced).

Relating marginal revenue to total revenue

You can get a handle on why the marginal revenue curve falls so quickly if you first examine total revenue, or *TR*. The total revenue that the monopoly can get is simply the output it produces times the price at which it can sell its output. That is, $TR = p \times q$. However, the price at which a monopoly can sell depends on how much it produces.

The relationship between output produced and the price at which it can be sold depends on the demand curve. For example, consider point *A* on the demand curve in Figure 12-1. At that point, one unit is being produced, and it can be sold for £9. Consequently, the total revenue at that point is £9. Next, look at point *B*, at which two units of output are being sold. At that output level, each unit can be sold for £8. Consequently, total revenue is $£8 \times 2 = £16$. And at point *C*, where three units can be sold for £7 each, total revenue is £21.

The important thing to notice is how total revenue changes as you move from *A* to *B* to *C* and output increases from one to two to three units. Total revenue goes from £9 to £16 to £21. Obviously, total revenue increases.

But look more deeply. Moving from *A* to *B*, *TR* increases by £7 (from £9 to £16). But moving from *B* to *C*, it increases by only £5 (from £16 to £21). Each successive increase in total revenue is smaller than the previous increase.

Increasing production, decreasing marginal revenue

Because marginal revenue is defined as the change in total revenue that happens as you increase production by one unit, the phenomenon we describe in the previous section is the same thing as saying that marginal revenue declines as the monopoly increases production.

If you look at Table 12-1, you can see that marginal revenue continues to fall for each successive unit. In fact, it becomes negative for all units after the

fifth. You can see why by looking at points *G* and *H* in Figure 12-1 as examples. At point *G*, the monopoly can sell seven units of output for £3 each, making for a total revenue of £21. But if it increases output to eight units at point *H*, the monopoly can sell these units for only £2 each, implying a total revenue of £16.

Increasing output from seven units to eight units means decreasing total revenue from £21 to £16. That's the same thing as saying that marginal revenue is *negative* £5 as you move from seven to eight units of output.

Sliding down the demand curve: Higher output, lower prices

Marginal revenue keeps declining and even becomes negative because the demand curve slopes downward, meaning that the only way to get people to buy more stuff is to offer them a lower price. You have to offer them a lower price not just on additional units, but on all previous units as well.

In other words, if the monopoly firm wants to sell only one unit (see point *A*), it can get £9 for that unit. But if the monopoly wants to sell two units (see point *B*), it has to lower the price down to £8 per unit for *both* the first unit *and* the second unit.

Because total revenue equals price times quantity ($TR = p \times q$), you can see that the monopoly faces a trade-off as it increases production and slides down the demand curve. As the firm produces more, q obviously goes up, but p must fall. What happens to TR depends on whether the increases in q (output effects) are bigger than the decreases in p (price effects).

You can see from Table 12-1 that as the monopoly increases production through the first four units, total revenue keeps increasing, meaning that the gains from selling more units more than offset the declines from getting less money per unit. At an output of five units, the two effects cancel each other out. And for higher outputs, total revenue falls because the negative effect of less money per unit overwhelms the positive effect of selling more units.

Because marginal revenue tells you how total revenue changes as you increase output, the changes in TR caused by increasing output show up in MR as well. If you look at Figure 12-1, you can see that MR is always declining. That's because the negative price effect of getting less per unit keeps getting stronger and stronger relative to the positive quantity effect of selling more units.

Pay attention, here comes the science! As we note earlier in the chapter, for straight-line demand curves like the one you see in Figure 12-1, the MR curve is a straight line with twice as steep a slope as the demand curve. If you know calculus, you can prove that the MR curve falls twice as fast as the

demand curve by taking the equation of the demand curve shown in Figure 12-1, $p = 10 - q$; substituting it into the total revenue equation, $TR = p \times q$; and then taking the first derivative with respect to output, q . Because marginal revenue is dTR/dq , you find that $MR = 10 - 2q$, meaning that MR has the same vertical intercept as the demand curve but twice as steep a slope.

Now that you've seen the marginal revenue situation facing a monopoly, you can combine it with the firm's marginal cost curve to figure out the profit-maximising output level. As we show you in the next section, this level is less than that chosen by a competitive firm – a behaviour that leads to social harm, which can be quantified using the method of deadweight losses that we explain in detail in Chapter 11.

Choosing an output level to maximise profits



A monopoly is no different to a competitive firm when it comes to the costs of producing output. Just like a competitive firm, a monopoly has fixed costs, variable costs and marginal costs (see Chapter 10). More importantly, these costs all behave in exactly the same way whether a firm is competitive or a monopoly. This situation means you can use costs to help analyse the decision-making process of a monopoly in the same way that you use them to analyse the decision-making process of competitive firms.

The key difference, however, is that the monopoly faces a downward-sloping marginal revenue curve. As we show in the next section, this factor causes a profit-maximising monopoly to produce less output than a profit-maximising competitive firm.

Setting $MR = MC$ for a monopoly

The monopoly goes about maximising profits in much the same way as a competitive firm. To see this, take a look at Figure 12-2, in which we draw a monopoly's average total cost (ATC) and marginal cost (MC) curves on the same graph as the monopoly's demand curve and marginal revenue (MR) curve.

As we explain in Chapter 10, for every output level, q , the ATC curve gives the average total cost per unit of producing q units of output. This curve is U-shaped because average total costs first fall due to increasing returns and then increase due to diminishing returns. The marginal cost curve gives the cost of producing one more unit of output; that is, it tells you how much total costs rise if you increase output by one unit.

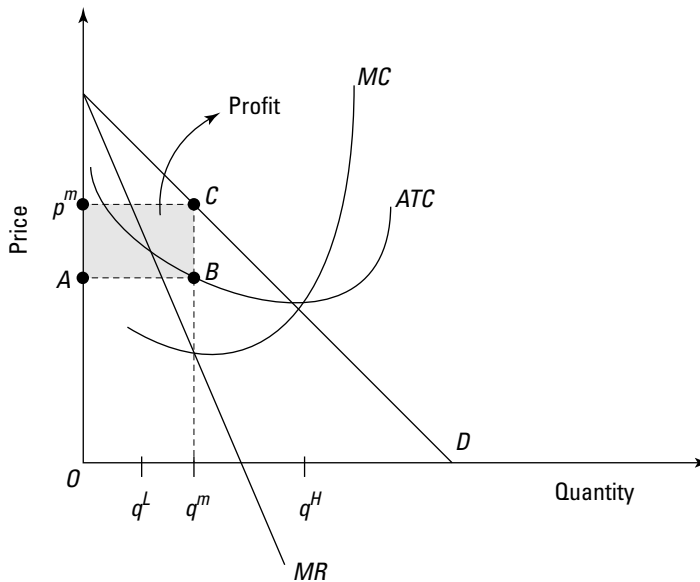


Figure 12-2:
A monopoly produces where its MC curve crosses the MR curve.

As Figure 12-2 illustrates, the profit-maximising monopoly's optimal output level, q^m , is determined by where the MR and MC curves cross. As with a competitive firm, choosing to produce where marginal revenues equal marginal costs ($MR = MC$) maximises profits or minimises losses, depending on whether demand is strong enough for the firm to be able to make a profit (see Chapter 10).

The reason that q^m is optimal can be seen by looking at two different output levels, q^L and q^H , where L stands for low and H stands for high:

- ✓ **Low output:** At output level q^L , you can go up vertically to see that MR at that output exceeds MC , meaning that if you produce and sell that unit, it brings in more in revenue than it costs to produce. Clearly, this unit is a good one to produce. Because a similar relationship holds true for all output levels less than q^m , the monopoly should keep increasing output until it reaches q^m .
- ✓ **High output:** On the other hand, the monopoly doesn't want to increase output beyond q^m . To see why, examine output level q^H . At that output level, marginal costs are much bigger than marginal revenues, meaning that if you produce that unit of output, the cost of producing it exceeds the money you can get selling it. In other words, if you produce that unit, you lose money.

So, as you can see, the monopoly wants to produce exactly q^m units, because for all units up to q^m , marginal revenues exceed marginal costs, meaning that you receive more money selling such units than you spend producing them.

Working out what price to charge

To figure out what the price of each unit of output needs to be, use the demand curve. Move up vertically from the monopoly's profit-maximising output level to the demand curve and then head sideways. In Figure 12-2, you can see that at output level q^m the monopoly can charge price p^m .

Eyeing the monopoly's profit

In Figure 12-2, the profit that the monopoly makes is shown by the shaded rectangle with corners at A , p^m , C and B . As we discuss in Chapter 10, such profit rectangles are derived by comparing the two rectangles that give, respectively, total revenues and total costs.

The basic trick is to remember that the area of a rectangle is defined as a product – the product of its length times its width. For the monopoly that is maximising profits by producing q^m units and selling them for p^m pounds, total revenue is price times quantity ($TR = p^m \times q^m$). Consequently, total revenue is the area of the rectangle whose length is equal to the price and whose width is equal to the quantity. That is, TR is the area of the rectangle that has corners O , p^m , C and q^m .

You can derive a total cost rectangle by first realising that total costs are also a product – a product of the average cost per unit times the number of units. If you go up vertically from point q^m until you hit the ATC curve, you get to point B . The vertical distance up to point B gives the average cost per unit of producing output q^m . So if you multiply that amount by the output q^m , you get total costs. Geometrically, therefore, total costs are given by the rectangle whose corners are O , A , B and q^m .

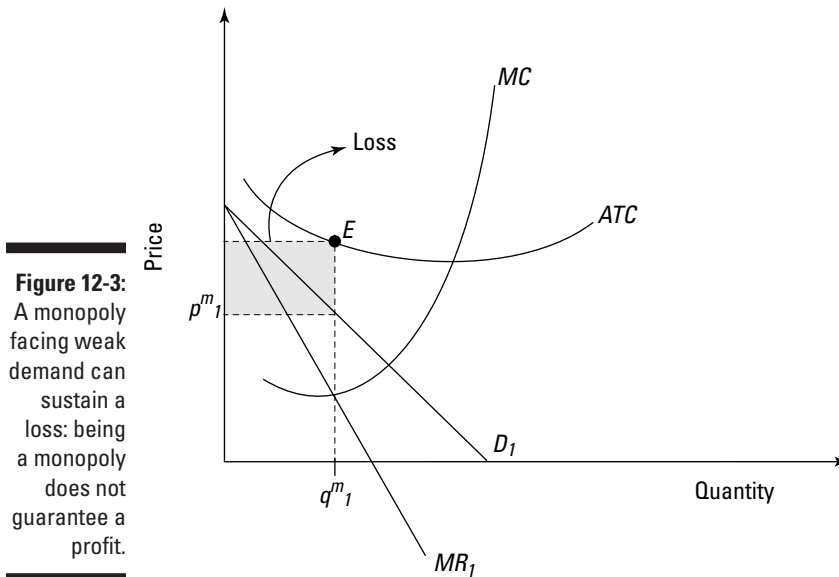
In Figure 12-2, the total revenue rectangle (O , p^m , C , q^m) is bigger than the total cost rectangle (O , A , B , q^m), meaning that the monopoly is earning a profit. That profit is given by the shaded rectangle whose points are A , p^m , C and B , which represents the difference in areas between the total revenue and total cost rectangles.



Understanding that monopoly doesn't guarantee profitability

Just because a firm has a monopoly doesn't mean that it's guaranteed a profit. If demand is too weak, prices are too low to make any money.

To see an example of this situation, look at Figure 12-3, where we draw a situation with very low demand. The new demand curve, D_1 , leads to a lower marginal revenue curve, MR_1 .



The monopoly again sets marginal revenue equal to marginal cost to find the optimal output level, q_1^m . But because of weaker demand, the monopoly operates at a loss represented by the area of the shaded rectangle.

One way to see that the shaded rectangle gives a loss is to compare the total revenue rectangle with the total cost rectangle, as we do for Figure 12-2 in the previous section. In this case, the total cost rectangle exceeds the total revenue rectangle by the amount of the shaded rectangle.

To understand where the loss comes from in a different way, compare the monopoly's average total cost per unit with the price per unit it gets when producing and selling at output level q_1^m . At that output level, the price per unit, p_1^m , is found by starting on the horizontal axis at q_1^m and then going up vertically to the demand curve. As you can see, you have to go up even farther to get to the ATC curve, meaning that the average total cost per unit to make q_1^m units exceeds the price per unit you get from selling these units. This fact implies that the firm loses money producing at output level q_1^m .

As we show in Chapter 10, a firm in such a situation can't do any better. That is, any other output level besides q_1^m produces an even bigger loss. If the monopoly can't work out a way to reduce costs or increase demand, it's quickly going bankrupt.

So keep in mind that even if you're the only seller in an industry, low demand may mean that you can't cover your production costs and make a profit.

Comparing Monopolies with Competitive Firms

So far in this chapter, we've examined how a monopoly acts in order to maximise its profits. We now want to compare a profit-maximising monopoly with a profit-maximising competitive firm. This comparison comes off very badly for the monopoly because, as we explain in Chapter 11, competitive firms deliver socially optimal output levels. Because monopolies always end up producing less than competitive firms, their output levels are always less than socially optimal.

Looking at output and price levels

Monopolies produce less than competitive firms because they have different marginal revenue curves. As we show earlier in the chapter, monopolies face downward-sloping marginal revenue curves. By contrast, competitive firms face horizontal marginal revenue curves.

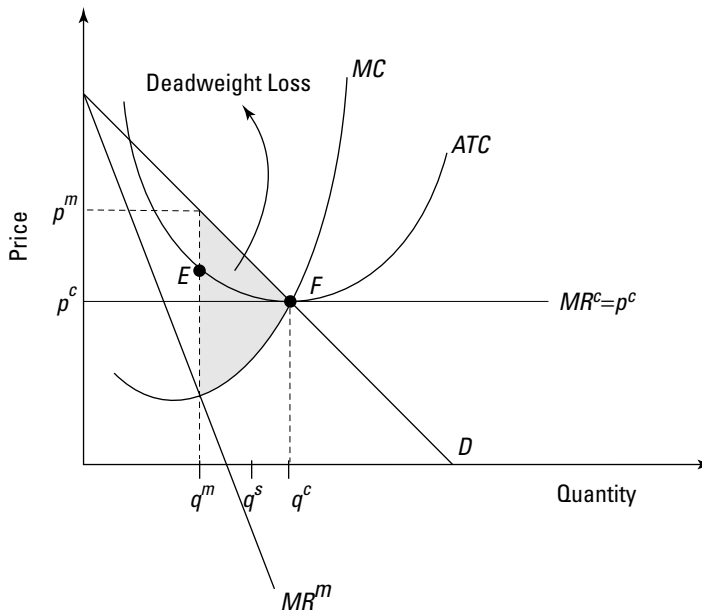
You can see the comparison in Figure 12-4, in which we draw both the downward-sloping marginal revenue curve of a monopoly, MR^m , and the horizontal marginal revenue curve of a competitive firm, MR^c . The graph also has an average total cost curve, ATC , as well as a marginal cost curve, MC .

Figure 12-4 assumes that the competitive firm and the monopoly have the same cost structure, which is why we show only one MC curve and one ATC curve. By assuming that both firms have the same cost structure, we can isolate the effect that the difference in marginal revenue curves has on each firm's output decisions.

Maximising profits for each firm

As we explain in Chapter 11, the marginal revenue curve for a competitive firm, MR^c , is a horizontal line set at the market price, p^c , because a competitive firm is such a small part of its industry that it can't affect the market price. As a result, it can sell as many or as few units as it wants at p^c , meaning that the marginal revenue it gets for every unit it chooses to produce is p^c . As we show in Figure 12-4, $MR^c = p^c$ for a competitive firm.

Figure 12-4: If a monopoly and a competitive firm have the same cost structure, the monopoly produces less, which causes a deadweight loss.



In addition, we show in Chapter 11 that market forces adjust supply and demand until the market price is equal to the minimum average total cost at which a firm can produce. Geometrically, this fact means that the horizontal $MR^c = p^c$ line just touches the bottom of the U-shaped ATC curve.

As we note earlier in the chapter, monopolies and competitive firms follow the same basic rule to maximise profits: they each produce where their marginal revenue curve intersects their marginal cost curve. But because they have different marginal revenue curves in Figure 12-4, they produce different outputs. The competitive firm produces q^c , whereas the monopoly produces q^m .

Understanding why the monopoly produces less

The competitive firm produces more than the monopoly, because the competitive firm doesn't have to worry about reducing its revenue per unit if it increases output. No matter how much it produces, the competitive firm always receives $MR^c = p^c$ on every unit sold because its output is too small relative to total output to affect the market price.

In contrast, the monopoly faces the market demand curve, meaning that every additional unit it sells lowers the price per unit it receives on all units sold. Geometrically, this fact implies the downward-sloping MR^m , which leads the monopoly to restrict output because it knows that the more it produces, the less money per unit it gets.

Because the monopoly restricts output compared with the competitive firm, the monopoly price, p^m , is also higher than the competitive price, p^c . This fact really irks consumers, but, as we note in the next section, the real harm comes from the reduction in output.

Deadweight losses: Quantifying the harm caused by monopolies

Monopolies cause harm because they reduce output below the socially optimal level produced by competitive firms. Take another look at Figure 12-4 and consider whether it would be good for society if all the units of output between the monopoly output level, q^m , and the competitive output level, q^c , were produced.

For example, look at unit q^s . At that level of output, the demand curve is above the marginal cost curve. This level implies that people are willing to pay more for that unit of output than it costs to make it. In other words, benefits exceed costs for that unit of output. Because this fact applies for all units between q^m and q^c , monopolies hurt society by failing to produce units of output for which benefits exceed costs.

The harm caused to society when the monopoly fails to produce output level q^s can be quantified by the vertical distance between the demand curve and the marginal revenue curve above output level q^s . That vertical distance is an amount that we can write down in money terms as the number of pounds by which benefits would exceed costs for that unit if it was produced and consumed.

If we go through the same exercise for each and every unit between q^m and q^c , we can total up the harm caused by the failure of the monopoly to produce those units. Graphically, the total harm measured in pounds is equal to the area of the shaded deadweight loss triangle in Figure 12-4.

As we explain in Chapter 11, you get this triangle by thinking of all the vertical distances between the demand curve and marginal cost curve that lie above all the units of output between q^m and q^c . All these little vertical distances shade in the deadweight loss triangle, which sums up the pound losses that result when the monopoly restricts output.

The deadweight loss triangle demonstrates that when monopolies restrict output in order to maximise their profits, they fail to produce units for which benefits exceed costs – and that harms society. The next section shows that the decision to restrict output also drives up production costs.

Focusing on efficiency

Another problem with monopolies is that they aren't efficient producers. To see that this is true, look again at Figure 12-4.

Competitive firms produce at output level q^c . If you go up vertically from that output level to the *ATC* curve, you end up at point *F*, which happens to lie at the very bottom of the U-shaped *ATC* curve. As we explain in Chapter 10, competition leads competitive firms to produce at the output level that puts them at the bottom of the U-shaped *ATC* curve.

That output level minimises production costs per unit of output, which you can see by comparing q^c with any other output level. Whether you produce more or less than q^c , average costs per unit are going to be higher due to the *U* shape of the *ATC* curve.

In particular, look at the monopoly output level, q^m . If you go up vertically from that output level to the *ATC* curve, you get to point *E*. Because the vertical distance between the horizontal axis and *E* is longer than the vertical distance between the horizontal axis and point *F*, you know for certain that total costs per unit when producing the monopoly output level, q^m , are higher than those when producing the competitive output level, q^c . Consequently, a monopoly firm produces output at a more costly output level than a competitive firm.



This bad result is yet another manifestation of the fact that monopolies face downward-sloping marginal revenue curves. A competitive firm has an incentive to increase output all the way to q^c because doing so lowers per-unit production costs and can thereby increase profits. The same incentive exists for a monopoly, but is more than offset by the reduction in revenue that happens if the monopoly firm increases its output. As a result, the monopoly's profits are maximised at q^m even though q^c is the lowest-cost output level.

Considering Examples of Good Monopolies

In this chapter we show you that compared to a competitive firm, a monopoly produces too little at too high a cost and turns around and sells it for too much money. Given these three bad things, you may simply want to say, 'Red card!' and get rid of monopolies altogether. But if you did, you'd be acting a bit too hastily. In some cases, the benefits of monopolies outweigh their costs. In other cases, the benefits outweigh costs in certain circumstances and in still other situations, society is still deciding whether the costs or benefits are the greater.

Encouraging innovation and investment with patents

The protection offered by patents is the most obvious case in which monopolies do society a lot of good. Patents give inventors the exclusive right to market their inventions for 20 years, after which time their inventions become public property. That is, patents given inventors the right to run a monopoly for 20 years.



Monopolies are vitally important in the context of innovation because without them inventors are unlikely to ever see any financial reward for their hard work: copycats are likely to steal their ideas and flood the market with rip-offs, thereby collapsing the price. Consequently, in a world without patents, far fewer people would bother to put in the time, effort and money required to come up with new inventions.

To remedy this situation, nations all over the world offer patent monopolies to inventors. The result is faster innovation, much more rapid economic growth and much faster increases in living standards.

However, those patents are limited so as to strike a balance between giving inventors incentives to invent and preventing people from holding permanent monopolies (for instance, you can't patent something obvious, or for which prior art exists).

Reducing annoyingly redundant competitors

Societies have also stepped in to create monopolies in situations where competition means annoying redundancies. Consider the following examples:

- ✓ **Rubbish collection:** Bin lorries are extremely loud and annoying. If one firm has a monopoly on collection, you have to endure a loud, annoying lorry only once per week. But if, say, seven different sets of bin men compete, you may have to endure a noisy lorry every day if you and six of your neighbours each choose to use a different firm that picks up on a different day of the week.
- ✓ **Cable television:** Only one cable TV provider exists for the whole of the UK. Think about the cost of laying wires to your home, and you can understand why. If ten different cable TV companies compete for your business, you'd need ten different sets of cable TV wires running underground – at much greater expense than running just one set of wires.
- ✓ **Natural gas:** Laying the pipes that deliver natural gas is expensive, and laying down multiple grids of gas pipe in one area would be wasteful.

As a result, most utilities and local services are supplied by a monopoly, whether state or private. Each company is given a monopoly and is then regulated to make sure that it doesn't exploit customers. For example, the supply of water in London (Thames Water) is a monopoly. (See the upcoming 'Regulating Monopolies' section.)

Keeping costs low with natural monopolies

Another area in which society may decide that a monopoly rather than competition is best is in the case of what economists call *natural monopoly industries*, or *natural monopolies*.



An industry is a natural monopoly if one large producer can produce output at a lower cost than many small producers. A good example is electricity distribution. The enormous fixed cost of setting up a national power grid means that there's no way of setting up a grid to serve a fraction of the market. In this case it has to be all or nothing.

Such an industry is called a *natural monopoly* because it naturally becomes dominated by a single low-cost producer. The perplexing problem here for policymakers is what to do with a natural monopoly.

On the one hand, everyone welcomes the fact that the grid serves the entire nation. But on the other hand, because there can be no economic competition, people now have to worry about the new monopoly charging high prices and producing less than the socially optimal output level.

These conflicting good and bad points typically mean that governments allow the natural monopoly to stay in business as the only firm in its industry, but at the same time they regulate it so that people don't have to worry about high prices or low output levels. By doing so, society gets the benefits that the most efficient production method generates without having to worry about the problems that may result if the monopoly was left unregulated.

Regulating Monopolies

Governments have to decide when to support and when to suppress monopolies. For example, patents support an inventor's monopoly right to produce and sell their invention for 20 years. After that, the production and sale of the invention is thrown open to competition.

In some monopolistic situations, various regulatory institutions have been developed to decide whether to destroy a monopoly by breaking it apart or

let it continue to be the only firm in its industry and regulate it. In this section, we present several of these regulatory schemes and explore what they do to improve the behaviour of monopolies.

Subsidising a monopoly to increase output

We establish in the ‘Comparing Monopolies with Competitive Firms’ section, earlier in this chapter, that a profit-maximising monopoly produces less than the socially optimal level. In particular, the profit-maximising monopoly in Figure 12-4 produces where its downward-sloping marginal revenue curve, MR^m , intersects its upward-sloping marginal cost curve, MC . This output level, q^m , is less than the socially optimal output level that would be produced by a competitive firm, q^c .

One way to get the monopoly to produce more is to subsidise its production costs so that the marginal cost curve in effect shifts down vertically. Doing so causes the marginal cost and marginal revenue curves to meet at a higher level of output. And if the subsidy is big enough, the monopoly can be induced to increase output all the way to q^c .

Some governments use this type of subsidy to get gas, electricity and phone companies to serve more people, especially poor people. If the monopoly firms’ costs of hooking up customers are subsidised, the firms are willing to hook up more customers than they would without the subsidy.

Some people object to subsidising a monopoly, so this sort of solution isn’t necessarily the most popular politically; but it is effective in increasing output.

Imposing minimum output requirements

Another way to get a monopoly to produce more is simply to order it to produce more. For example, in some places telephone companies are required to provide basic telephone service to everyone – even to people who can’t pay for it themselves. (The idea is to make sure that everyone is able to call for help in case of an emergency.) The same is often true of companies that provide heating in the winter; in some places, you can’t turn off someone’s heat for non-payment of bills.

Minimum output requirements can force a monopoly to produce the socially optimal output level. They are often politically popular because many people think of monopolies as evil and exploitative and don’t mind seeing them ordered to produce more.

Any forced increase in output also means a reduction in the monopoly's profit. Therefore, such programmes are also popular because many people consider a monopoly's profits to be ill-gotten given the fact that the firm doesn't have to compete to earn them.

Regulators have to be careful, however, not to bankrupt a monopoly when regulating. Depending on a monopoly's cost curves, forcing a monopoly to produce at an output level where it loses money is quite possible. Because regulators don't want to bankrupt monopolies and thereby deny consumers access to the goods or services they produce, regulators are careful to take a monopoly's cost structure into account when considering minimum output requirements.

Regulating monopoly pricing

Perhaps the most common way to regulate a monopoly is to set the price at which it can sell each and every unit of output that it produces. This approach works because it changes the monopoly firm's marginal revenue curve from sloping downward to being horizontal. Therefore, it eliminates the monopoly's usual problem that the more it sells, the less it can charge per unit.

However, as with quantity requirements, regulators have to pay close attention to a monopoly's cost structure when choosing the regulated price so that they don't bankrupt the monopoly.

To see the problem facing the regulator, consider the monopoly whose cost curves are given in Figure 12-5. Left unregulated, the monopoly chooses to produce the profit-maximising output level q^m , defined by where MR crosses MC . From the demand curve, you can see that it's able to charge price p^m per unit for that amount of output. (For more on the behaviour of an unregulated monopoly, see the earlier section 'Choosing an output level to maximise profits'.)

Next, think about how a regulator may want to modify the monopoly's behaviour. For example, a well-intentioned regulator may want to get the monopoly to produce every single unit of output for which benefits exceed costs. Looking at Figure 12-5, you can see that the regulator wants to get the monopoly to produce output level q^{mc} , defined by where the downward-sloping demand curve intersects the MC curve.

As we explain in the earlier section 'Deadweight losses: Quantifying the harm caused by monopolies', producing each unit up to and including q^{mc} is socially beneficial. That's true because, for each unit, what people want to pay to consume it (given by the vertical distance from the horizontal axis up to the demand curve) exceeds the marginal cost of producing it (given by the vertical distance from the horizontal axis up to the MC curve).

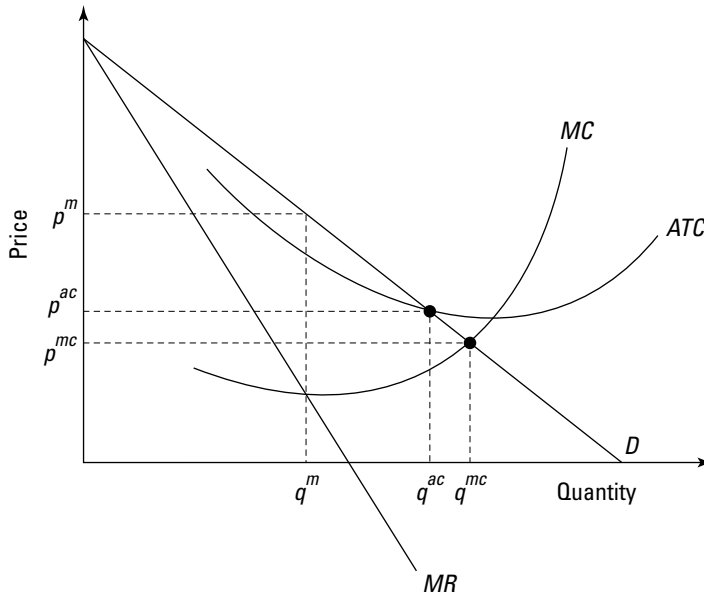


Figure 12-5: Regulating a monopoly using average cost pricing and marginal cost pricing.

As a result, the regulator sets the price at p^{mc} . At that price, the demand curve tells us that consumers want to purchase q^{mc} units of output. Better yet, the monopoly wants to supply that level of output because the marginal revenue from selling each unit exceeds the marginal cost of producing it.

However, a big problem exists with this policy given this particular monopoly's cost structure: the monopoly goes bankrupt, because at output level q^{mc} , the firm's total costs exceed its total revenues.

You can see this problem on a per-unit basis by noting that the average total cost per unit at output level q^{mc} (given by the vertical distance from the horizontal axis up to the ATC curve) is more than the regulator-imposed revenue of p^{mc} per unit. Because average total costs per unit exceed revenues per unit, the monopoly is going to be operating at a loss. If the regulator doesn't relent and allow a higher price, the monopoly eventually goes bankrupt, unless the government steps in to subsidise the firm by giving it a cash grant equal to the amount of its loss.



The method of regulation we just described is called *marginal cost pricing*, because the regulated price, p^{mc} , is set where the marginal cost curve crosses the demand curve. But because this method can cause a monopoly to lose money, a more common alternative is *average cost pricing*, which sets the regulated price where the average total cost curve (ATC) intersects the demand curve.

In Figure 12-5, a regulator using average cost pricing would set the price at p^{ac} . At that price, you can see from the demand curve that consumers demand q^{ac} units of output. The monopoly is happy to supply that output level because for each and every unit up to q^{ac} , marginal revenue (the regulated price per unit, p^{ac}) exceeds marginal cost – meaning that the monopoly gains financially by producing each and every one of these units.

The main benefit of this system is that you don't have to worry about the monopoly going bankrupt (or where to get the money to subsidise a monopoly that would go bankrupt under marginal cost pricing). Average cost pricing guarantees that the monopoly is going to break even.

You can see this fact by comparing the average total costs per unit at output level q^{ac} with the revenue per unit at that output level. You get the average total cost per unit by going up vertically until you hit the average total cost curve. Because that vertical distance is equal to the regulated price per unit, p^{ac} , you know that the average total costs per unit are equal to the regulated price per unit – and so the firm must be breaking even.

The downside to average cost pricing for this monopoly is that all the socially beneficial units between q^{ac} and q^{mc} don't get produced. On the other hand, the only way to keep this monopoly in business to produce those units if you imposed marginal cost pricing is to subsidise it. Using average cost pricing eliminates any worries associated with providing subsidies. In particular, you don't have to worry about any potential harm that you may cause when raising the taxes that have to be imposed somewhere else in the economy in order to subsidise the monopoly.

Breaking up a monopoly into several competing firms

Another solution to the problem of a monopoly is to destroy it by breaking it up into many competing firms. (It's fair to say that this is regarded as something of a nuclear option in competition law, so it happens rarely; in general other penalties are preferred.) The most famous use of this solution was the division of American Telephone and Telegraph Corporation (AT&T) into a bunch of smaller competitors in 1984. (Economists often like to compare this treatment of AT&T with the more lenient handling of British Telecom by UK authorities.)



Before 1984, AT&T was a nationwide monopoly. If you wanted to make a telephone call anywhere in the United States, you had to use AT&T because it was the only telephone company in the country. It was highly regulated, with both quantity requirements to provide everyone a phone and price requirements

that encouraged it to provide a high quantity of telecommunication services. But it was still a monopoly, and a judge ruled in 1984 that it should be broken up into numerous local firms in order to foster competition.

The policy change worked extremely well. A very competitive market for telephone services soon emerged between firms that had been part of AT&T. And more recently, the telephone service industry has become even more competitive due to the arrival of mobile phone companies, Internet telephony companies and even cable TV companies offering phone services. This robust competition eliminates the problems associated with monopolies and ensures that telecommunication services are provided at low cost and in large quantity.

Creating competition is also a handy way to deal with a monopoly because it eliminates the costs associated with having to continually monitor a regulated monopoly. As we explain in Chapter 11, competition gets you to the socially optimal output level without any sort of central control. That stands in stark contrast to regulated monopolies, which typically require expensive bureaucracies to develop and enforce laws and regulations.

Chapter 13

Oligopoly and Monopolistic Competition: Middle Grounds

In This Chapter

- ▶ Deciding whether to compete or collude in an oligopoly
- ▶ Understanding how colluding firms hurt consumers and society
- ▶ Using the Prisoner's Dilemma model
- ▶ Examining why some collusive pacts work whereas others don't
- ▶ Regulating firms so they can't collude
- ▶ Using product differentiation to elude perfect competition
- ▶ Limiting profits in monopolistic competition

Chapters 10 and 12 examine in detail the two most extreme forms that an industry can take: perfect competition, which features many small competitive firms, and monopoly, where only one firm exists (and hence no competition). This chapter concentrates on two interesting intermediate cases.

The first case is an *oligopoly*, an industry in which only a small number of firms operate – two, three or a handful. The word itself is Greek for ‘few sellers’. A diverse group of industries looks like this, including soft drinks, oil production and video game consoles. For example, Coke and Pepsi dominate the soft drink market, vastly outselling other carbonated beverages. Similarly, just three or four countries produce the majority of the world's oil. And just three companies produce and sell virtually all the video game consoles used in the world.

Oligopoly industries are interesting because, depending on specific circumstances, the firms can compete ruthlessly with each other or unite to behave almost exactly like a monopoly. This situation means that in some cases oligopolies can be left alone because competition ensures that they produce socially optimal output levels, whereas in other cases government regulation is

needed to prevent them from acting like monopolies and behaving in socially undesirable ways.

The second type of intermediate industry is one where you find *monopolistic competition* – a sort of hybrid between perfect competition and monopoly. The key thing that sets firms in this type of industry apart from firms in a perfectly competitive industry is *product differentiation* – the fact that each firm produces a slightly different product than the others. A very good example of an industry that is only ever monopolistically competitive is television, where the competition for viewers is about which strain of programming is being shown.

This chapter starts with a detailed look at oligopolies and the decisions that firms in this type of industry have to make. We then move on to monopolistic competition and show why product differentiation doesn't necessarily translate into tidy profits.

Choosing to Compete or Collude

In industries where only a few firms operate, the firms have a choice about whether to compete or co-operate. This situation is very different from perfect competition, which we discuss in Chapter 11.

In perfectly competitive markets, the large number of firms and the fact that each firm is only a small part of the market means that their individual outputs don't have any effect on the market price. As a result, competitive firms just take the market price as given and adjust their output levels accordingly to make as large a profit as possible.

Realising that oligopoly firms interact strategically



In a market in which only a few sellers operate, each one produces enough of the total output to be able to affect the market price. For example, there are two major producers of cola-flavoured fizzy drinks: Coke and Pepsi. These two corporations produce such large fractions of the total output that if one were suddenly to increase supply, the market price of cola-flavoured fizzy drinks would drop dramatically. An increase in output made by one company causes the price to decrease for other companies in the market as well.

In other words, if Pepsi produces twice as much of its product and literally floods the market, the price of Pepsi drops dramatically. But because most people aren't 100 per cent loyal to one brand or the other, if the price of Pepsi drops dramatically, a lot of regular Coke drinkers are going to switch brands and drink Pepsi. In turn, the price of Coke drops too.

Pepsi and Coke are involved in a situation where each of their supply decisions affects not only their own sales, but also those of their competitor. Economists refer to such situations as *strategic situations*, because the firms involved have to decide what type of strategy to pursue. In particular, they have to decide whether to compete or collude:

- ✓ If they collude, they jointly cut back on production in order to drive up prices and increase their profits.
- ✓ If they compete, they both try to increase production in order to undercut each other on price and capture as many customers as possible.

Comparing the outcomes of competition and collusion

These two strategies, compete or collude, lead to hugely different outcomes for both producers and consumers:

- ✓ For producers, collusion is better than competition because it leads to profits that last as long as the firms keep colluding.
- ✓ For consumers, collusion is worse than competition because it leads to higher prices and lower output.

Seeing these results, you may assume that government intervention is called for in order to protect consumers from collusion. But such intervention is needed only if firms actually collude.

A fascinating thing about the real world is that collusion doesn't happen in a lot of industries where you'd expect it. For example, Coke and Pepsi are fierce competitors that spend hundreds of millions of pounds a year on advertising to try to steal each other's customers.

Similarly, most territories have only a handful of competing mobile phone companies. But instead of colluding, they compete so ruthlessly that many of them are constantly flirting with bankruptcy. Other industries with few competitors that are fiercely competitive include low-cost airlines, where market shake-outs have been common.

The big question that economists have to answer is, ‘Why do we see so little collusion in industries where you would expect more of it?’ The next few sections show you how economists respond.

Cartel Behaviour: Trying to Imitate Monopolists



A group of firms that colludes and acts as a single co-ordinated whole is known as a *cartel*. Because a cartel acts essentially as one gigantic firm, it effectively turns a bunch of individual firms into a single big monopoly.

This fact makes understanding the profit-maximising behaviour of a cartel easy, because that behaviour is just like that of a monopoly. In fact, you can see what a cartel wants to do by looking at the figures in Chapter 12, which illustrate what a monopoly likes to do.

In particular, a profit-maximising cartel chooses to produce the monopoly’s profit-maximising output level of q^m units shown in Figure 12-2. Producing that output level maximises the cartel’s collective profit, which is shown as the shaded area of Figure 12-2. And better yet for the cartel, that monopoly profit persists as long as the participating companies keep co-operating and producing a combined output of q^m .

Co-ordinating a cartel is hard work

Unfortunately for the firms in the cartel, getting all the firms to co-ordinate so that they are collectively producing the monopoly output level, q^m , is often very difficult. To get the individual firms to co-operate and produce exactly q^m units of combined output, you have to get them to agree about two related things:

- ✓ **How to share the profits:** Obviously, every firm wants as large a share as possible.
- ✓ **How to set output quotas:** The firms must agree, and abide by, how much of the total output (q^m) each firm produces. Each firm is constantly tempted to produce more than its quota because doing so would bring higher revenues.

Examining OPEC to see the difficulties of collusion



The difficulties of meeting these two requirements are illustrated by the OPEC oil cartel. OPEC stands for *Organisation of the Petroleum Exporting Countries*. Although OPEC is a dull name, the group is very lively, and includes Saudi Arabia, Iraq, Venezuela, Nigeria, Kuwait, Indonesia and several other key oil-exporting nations.

Together, these nations control the vast majority of the world's oil reserves. They sell the right to extract oil to the major oil companies who in turn sell it on to the consumer. The importance of reserves means that they make up an oligopoly industry with only a few participants. Because only a few players exist, they have a chance to form a cartel and try to produce the monopoly output and make monopoly profits. Do they succeed?

On the whole, no. We say 'on the whole' because although they do negotiate agreements about oil production, these agreements are constantly broken. For example, suppose that the monopoly output level that maximises OPEC's collective profit is 20 million barrels per day, and at that output level the price of oil is US\$60 per barrel.

To achieve that combined output, OPEC has to agree on each country's production quota. For example, Saudi Arabia may get to produce 4 million barrels per day while Venezuela may have a quota of only 2 million barrels per day, leaving the other 14 million barrels per day to be split up among the other member countries.

Unfortunately for OPEC, enforcing the quotas is impossible. In particular, no way exists to stop Venezuela from pumping more than its 2 million barrels per day and selling the excess onto the world oil markets. Nearly all the OPEC countries cheat and overproduce.

They do so because the high price of oil is just too tempting. For example, if all the other countries obey the agreement and drive up the price of oil, Venezuela finds it very tempting to produce more than its quota because each additional barrel it produces brings in lots of money.

Unfortunately, this same temptation faces each country, so they nearly all overproduce their quotas. All their cheating causes an increase in supply, which floods the market and reduces the price to well below what the countries would receive if they stuck to their respective quotas.

Put slightly differently, cartels have self-destructing incentives: they do work and create monopoly profits, but they also create temptations for cartel members to cheat. In the case of OPEC, these temptations are so strong that OPEC has only occasionally been able to act as an effective cartel. (And this occasional success is largely down to Saudi Arabia's dominant position within the cartel – that is, when other members cheat on the arrangement, Saudi Arabia has been able to use its greater capacity to threaten to dump its output on the market, forcing the cheats back to the negotiating table.)

In the next section, we go over a game theory model called the *Prisoner's Dilemma*, which gets to the heart of why firms in cartels cheat and why, in many cases, stopping them from doing so is nearly impossible.

Understanding the Prisoner's Dilemma Model

The behaviour of cartels and their incentive to cheat is best understood by applying to cartels a very famous game theory model known as the *Prisoner's Dilemma*.

Game theory is a branch of mathematics that studies how people behave in strategic situations – situations in which their actions or anticipated actions are taken account of by other people who then modify their own actions accordingly. If you've seen the film *A Beautiful Mind*, you've come across some of the models used.

For example, chess and checkers are strategic situations because what you do on your current move changes what your opponent does in subsequent moves. Even more important, what you *think* your opponent is going to do in response to each of the moves that you may make right now helps you to choose the best thing to do. That is, you take into account how your opponent is going to react to each of your possible actions before you decide the best thing to do.

Cartels are strategic situations too, because each firm has to take into account what it thinks all the other firms are going to do before deciding what it needs to do. Consequently, game theory models are the best way to understand the motivations and temptations that guide the behaviour of cartel members.

To keep matters simple, consider a *duopoly*, an industry in which only two firms exist. The two firms have the opportunity to form a cartel, act as a monopoly would and generate a monopoly profit which can then be shared.

But do they? That all depends on whether either firm (or both) is going to cheat on the cartel agreement.

The best way to understand each firm's temptations is to first study the Prisoner's Dilemma, a game in which two criminal partners have to individually decide whether or not to cheat on an agreement they had previously made with each other to remain silent and not talk to the police about their illicit activities.

Fleshing out the Prisoner's Dilemma



Imagine that two criminals named Reggie and Ronnie have just robbed a bank. The police know this, but don't have any hard evidence against them. Rather, their only way of getting a conviction is to get one or both of the bank robbers to confess to the crime and give evidence against the other.

Fortunately for the police, they do have some leverage because they managed to catch Reggie and Ronnie committing other unrelated, minor crimes. These other crimes carry with them a one-year jail sentence. The police are hoping to use the threat of a year in jail to get one or both of the bank robbers to implicate his partner in exchange for immunity from prosecution.

Reggie and Ronnie had previously made a pact not to rat on each other, but the police are willing to find out what happens when push comes to shove.

Comparing the pay-offs of confessing or remaining silent

Following standard procedures, the police separate Reggie and Ronnie, questioning them in separate interrogation rooms, which prevents Reggie and Ronnie from communicating with one another, reduces their ability to renew their commitments and stops them from getting any idea about their partner in crime's next move. The police offer each of them the chance to give evidence against the other in exchange for immunity.

The problem for each man is that what happens to him depends not only on what he does, but also on what his partner does. Each man can trade a confession for immunity, but he gets the deal only if his partner doesn't confess at the same time in the other interrogation room.

Four outcomes are possible:

- ✓ If the men both keep their pact not to talk and neither confesses to robbing banks, each man gets only a year in jail for the minor offence.
- ✓ If Reggie confesses and agrees to give evidence against Ronnie while Ronnie remains silent, Reggie goes free because he co-operated with police, but Ronnie gets ten years for bank robbery.

- ✔ If Ronnie confesses and gives evidence while Reggie remains silent, Ronnie goes free and Reggie goes to jail for ten years.
- ✔ If both men admit to the crime, they each get five years in jail. Why five years? If both confess, the police don't need to make such a generous deal; they don't need to give either man immunity in order to get evidence against the other. On the other hand, the police want to give each criminal an incentive to confess, so they send each man to jail for only five years rather than the ten years he'd get if he remained silent while his partner gave evidence.

Putting the pay-offs into a matrix for easy comparison

Figure 13-1 contains a pay-off matrix. The figure illustrates the outcomes in terms of jail time that each bank robber receives depending on the decision that each man makes about whether to remain silent or confess.

		Reggie's Choices	
		Confess	Remain Silent
Ronnie's Choices	Confess	Reggie gets 5 years Ronnie gets 5 years	Reggie gets 10 years Ronnie gets 0 years
	Remain Silent	Reggie gets 0 years Ronnie gets 10 years	Reggie gets 1 year Ronnie gets 1 year

Figure 13-1: The pay-off matrix of jail times facing Reggie and Ronnie.

The columns give each of Reggie's options, confess or remain silent. The rows give each of Ronnie's options, which are the same. Each of the four rectangles in the grid shows the jail times that result from each of the four possible combinations of their individual decisions about whether or not to confess.

For instance, the upper-left rectangle represents what happens if both of them confess. The rectangle is divided diagonally in half, with Reggie's pay-off of five years in jail given in the upper shaded triangle and Ronnie's pay-off of five years given in the lower triangle. Similarly, the upper-right rectangle gives each of their pay-offs if Reggie remains silent while Ronnie confesses:

Reggie gets ten years in jail while Ronnie gets zero years because he gave evidence against Reggie.

Determining the dominant strategy for each prisoner

The Prisoner's Dilemma is famous because the way the police have set up the potential pay-offs means that each criminal has an incentive to confess – no matter what the other criminal does.

For example, concentrate on Ronnie. Should he confess or remain silent? Well, first examine which option is best for him if his partner in the other interrogation room confesses. Looking at the left column of pay-offs, you can see from the upper-left rectangle that if Ronnie confesses while Reggie confesses, Ronnie gets five years. On the other hand, the bottom-left rectangle tells you that if Ronnie remains silent while Reggie confesses, Ronnie gets ten years.

Clearly, the best thing for Ronnie to do if Reggie confesses is also to confess. But consider whether Ronnie should confess or remain silent while Reggie remains silent in the other interrogation room. Begin with the upper-right rectangle, which shows that if Ronnie confesses while Reggie remains silent, Ronnie gets zero years in jail. In contrast, the bottom-right rectangle tells you that if Ronnie remains silent while Reggie is also silent, Ronnie gets one year in jail. Clearly, if Reggie remains silent, the best thing for Ronnie to do is confess and get zero years in jail rather than one year.

In other words, the best bet for Ronnie is *always* to confess. If Ronnie confesses when Reggie confesses, Ronnie gets five years rather than ten. And if Ronnie confesses when Reggie remains silent, Ronnie gets zero years rather than one. So Ronnie should *always* confess no matter what Reggie is saying or not saying to the police in the other room.

Because the pay-offs of confessing are always better for Ronnie than the pay-offs of not confessing, game theorists refer to confessing as being Ronnie's *dominant strategy*, by which they mean superior strategy given the set of pay-offs.

If you go through the pay-offs from Reggie's perspective, you find that confessing is also Reggie's dominant strategy, because no matter what Ronnie is doing, the pay-offs to Reggie if he confesses are always better than those from remaining silent.

Confessing is thus a dominant strategy for both players, meaning that you should expect both of them to confess separately. If they do so, they end up in the upper-left box of the pay-off matrix, where they both get five years in jail.

Realising that the dominant strategy leads to a lousy outcome for both players

The police, of course, want each criminal to confess separately and go to jail for five years, which is why they keep the prisoners apart and set up the pay-offs in the way they do.

Reggie and Ronnie previously promised each other not to talk to the police, but pay-offs are structured so that confessing is the dominant strategy, which puts them in a bind. Each man keeps his promise and risks huge jail time if his partner confesses or breaks his promise in an effort to reduce his own potential jail time. This hard choice is why this situation is known as the Prisoner's Dilemma.

Typically, both men go with their dominant strategy and confess. But because both separately decide to confess, they each end up getting five years in jail – a much worse outcome than if they both kept their promise to remain silent and both went to jail for only a year. The logic of the dominant strategy is so compelling, though, that they each break the agreement and end up going to jail for five years rather than one.

As we show you later in the chapter (see the section 'Applying the Prisoner's Dilemma to Cartels'), cartel members also face a Prisoner's Dilemma because they must decide whether to obey the cartel agreement (to reduce output to the monopoly level) or to cheat and overproduce. As we show, the temptation for cartel members to overproduce and break their cartel output agreement is just as strong as the temptation for prisoners to confess and break their agreement not to talk to the police.

Using omerta to resolve the Prisoner's Dilemma

So how do you get around the problem? Organised crime has found one answer, which is to create a structure to enforce the bargains made by the cartel in an effective if bloodthirsty fashion.

The system is called *omerta*, which is Sicilian for 'silence'. Basically, what the Mafia does is change the pay-offs to the Prisoner's Dilemma so that the dominant strategy switches from confessing to remaining silent. The mob does this by explaining to their criminal members that if anybody talks to the police and confesses to anything or implicates anyone else, they're going to die.

This threat totally rearranges the pay-offs to the Prisoner’s Dilemma. Instead of just comparing jail times, as in Figure 13-1, prisoners now have to factor in death, as in Figure 13-2. If you look at Figure 13-2 carefully, you find that the dominant strategy for both players is now to remain silent because if either talks, the Mafia hunts him down and kills him no matter what the other man does. The result is that both Reggie and Ronnie go to jail for only one year each, because they both keep their mouths shut.

Figure 13-2:
The dominant strategy in the Prisoner’s Dilemma changes when death becomes the pay-off to confessing.

		Reggie’s Choices	
		Confess	Remain Silent
Ronnie’s Choices	Confess	Reggie dies Ronnie dies	Reggie gets 10 years Ronnie dies
	Remain Silent	Reggie dies Ronnie gets 10 years	Reggie gets 1 year Ronnie gets 1 year

Paradoxically, the death threat benefits the two criminals. Even though the threat of death is scary for both Reggie and Ronnie, it actually serves their individual interests because it means that they go to jail for only one year rather than five. If you’re going to be a criminal, you want to be part of a criminal organisation that has enough power to bully you around and keep you from renegeing on your agreements with your fellow criminals.

This idea of a strong enforcer is crucial. The following sections show that one way to get a bunch of firms to stick to a cartel agreement is if the cartel sets up a strong threat against cheaters in the same way that the Mafia threatens to kill anyone who breaks with the *omerta* agreement to never talk to the police. (We ought to acknowledge that it’s not a strategy a company can generally employ.)

Incidentally, the *omerta* system is effective not just because it is an external enforcement mechanism, but also, and more importantly, because of its effect on individual incentives. Notice in the example above that the carrot

accompanying the stick of punishment is that both men get out much more quickly (and probably spend the intervening year perfecting their marinara sauce amongst friendly fellow mob members). Arguably, the effect on individual incentives can make crime syndicates rise or fall.

Applying the Prisoner's Dilemma to Cartels



To see how the Prisoner's Dilemma applies to cartels, imagine a duopoly industry – an industry in which only two firms exist. Suppose the industry in question is the snack market in a town, and the only two makers of snacks are Jazy Drops, Ltd. and Fruity Rock, Inc.

The firms can compete against each other aggressively or restrict supply to keep prices high and make big monopoly profits. To keep things simple, imagine that the managers of both firms think of their situation in terms of two prices that they can charge, £3 per bag or £2 per bag.

Figure 13-3 shows a pay-off matrix for both firms depending on the choices each makes about which price to charge. The pay-offs are given in terms of each firm's profits per day, with those of Fruity Rock given in the shaded areas.

Fruity Rock's Choices

		Fruity Rock's Choices	
		Charge £3	Charge £2
Jazy Drops's Choices	Charge £3	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;">£1,000 per day for Jazy Drops</div> <div style="width: 10%; background-color: #cccccc; text-align: center;">£1,000 per day for Fruity Rock</div> <div style="width: 45%;">£500 per day for Jazy Drops</div> </div>	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;">£500 per day for Jazy Drops</div> <div style="width: 10%; background-color: #cccccc; text-align: center;">£2,000 per day for Fruity Rock</div> <div style="width: 45%;">£500 per day for Jazy Drops</div> </div>
	Charge £2	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;">£2,000 per day for Jazy Drops</div> <div style="width: 10%; background-color: #cccccc; text-align: center;">£500 per day for Fruity Rock</div> <div style="width: 45%;">£800 per day for Jazy Drops</div> </div>	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;">£800 per day for Jazy Drops</div> <div style="width: 10%; background-color: #cccccc; text-align: center;">£800 per day for Fruity Rock</div> <div style="width: 45%;">£800 per day for Jazy Drops</div> </div>

Figure 13-3: Each firm's profit depends not only on its own pricing decision, but also on that of the other firm.

For example, in the upper-left rectangle, where both firms charge £3 per bag, they each earn a daily profit of £1,000. In contrast, if they both charge £2 per bag, their individual daily profits fall to only £800, as you can see in the bottom-right rectangle.

Obviously, if these two firms collude and charge £3 rather than £2 per bag, they can raise profits by £200 per day each. The joint monopoly profit that they can earn totals £400 per day (£200 each). Their problem is to form a cartel that does, in fact, get them this monopoly profit.

Such a cartel is hard to maintain, however, because the temptation to cheat is ever-present. For example, the other two rectangles (the upper-right and bottom-left) show what happens if one firm charges £3 while the other charges £2. The firm charging £2 steals business away from the £3 firm and earns a much higher profit. Indeed, the firm charging £2 per day earns a profit of £2,000 per day while the firm charging £3 per day earns a profit of only £500, because this firm is losing business to its more cheaply priced competitor.

If you look at each firm's incentives, you notice that the dominant strategy for each firm is to charge the lower price, £2 per bag. To see this, look at Jazzy Drops's pay-offs. If Fruity Rock charges £3 per bag, the better thing for Jazzy Drops to do is to charge £2 per bag. You can see this by comparing Jazzy Drops's pay-off of £1,000 per day in profits in the upper-left rectangle with its pay-off of £2,000 per day in the bottom-left rectangle.

Similarly, if you look at the upper-right and bottom-right rectangles, you see that the best thing for Jazzy Drops to do if Fruity Rock charges £2 per bag is to also charge £2 per bag, because Jazzy Drops's profit is £500 per day if it charges £3 per bag but £800 per day if it charges £2 per bag.

All this means that no matter what Fruity Rock decides to charge, Jazzy Drops is always better off charging only £2 per bag.

If you go through the rectangles, you find that charging £2 per bag is also Fruity Rock's dominant strategy, no matter what Jazzy Drops charges. The result is that both firms always decide to charge £2 per bag – and they lose out on their chance to join forces, reduce output, drive up prices and earn monopoly profits.

If both firms somehow figured out a way to commit to selling at £3 per bag, they would end up in the upper-left rectangle and earn £1,000 per day each in profits. But without a way to commit, they're each going to follow their dominant strategy, charge £2 per bag and end up in the bottom-right rectangle earning only £800 per day in profits. By failing to work together, they each lose £200 per day in profits.



Of course, this situation is great for the town's consumers, who'd much rather pay only £2 per bag for their snacks. So keep in mind that the dominant strategy of charging the lower price works toward the benefit of consumers and society at large. This fact is why society often doesn't have to bother regulating oligopoly industries. Thanks to the Prisoner's Dilemma, cartels very often fail to raise prices.

Seeing that OPEC is trapped in a Prisoner's Dilemma

The basic version of the Prisoner's Dilemma that we show you in previous sections is set up for just two people or two firms. But mathematicians have developed more advanced versions of the Prisoner's Dilemma that can be used to analyse the behaviour of larger numbers of participants. These models are invaluable for understanding oligopoly industries with several firms and the incentives faced by firms in such industries when they try to form cartels. The basic conclusion of these multi-firm models is that the dominant strategy is usually to cheat on cartel agreements.

This result goes a long way toward explaining why the OPEC oil cartel has a hard time when it tries to achieve its goal of raising oil prices by reducing oil production. Quite simply, cheating on OPEC cartel agreements is a dominant strategy for OPEC member countries.

To see how this works, you have to first understand that OPEC has meetings where it decides how much total oil is to be produced and what fraction of that overall production is to be done by each country. At the meetings, each country is given a *quota* – a maximum amount supposedly to produce. For example, Saudi Arabia may be given a quota of 10 million barrels per day, while Venezuela may be given a quota of 1 million barrels per day.

The problems start after the meetings when all the oil ministers go home. Each country realises that producing more than its quota is the best strategy no matter what the other countries do. For example, Venezuela is better off producing more than its 1 million barrels a day quota no matter what the other countries do.

- ✓ If the other countries obey their quotas, Venezuela is better off producing more than its quota because it can sell lots of oil at a high price. (The high price is caused by the fact that the other countries are obeying their quotas.)
- ✓ If the other countries break their quotas and overproduce, the price of oil is low, meaning that Venezuela should also overproduce its quota. No reason exists to obey the quota if prices are low due to everyone else's cheating.

Because each country faces the same temptation to overproduce its quota, the OPEC cartel doesn't typically work very well. Overproducing is a dominant strategy and is simply too tempting to resist given the pay-offs.

Using an enforcer to help OPEC members stick to quotas

In the earlier section 'Using omerta to resolve the Prisoner's Dilemma', we explain how the Mafia uses death threats to get its members to never speak to the police. The death threats change the pay-offs so much that the dominant strategy in the Prisoner's Dilemma switches from confessing to remaining silent.

OPEC would benefit by threatening its members when they violate their quotas. Because the member countries are sovereign nations, death threats aren't an option. Instead, Saudi Arabia sometimes tries to provide an economic threat against quota violators.

This economic threat comes in the form of super low oil prices. Saudi Arabia is in the best position to make such a threat for two reasons:

- ✓ **Saudi Arabia is the world's largest oil producer.** It produces around 25 per cent of the world's oil.
- ✓ **Saudi Arabia is the world's lowest cost oil producer.** It can produce profitably even if the price of oil falls down to £2 per barrel. (At the time of writing other countries need a price of at least £6 per barrel to break even.)

These two facts mean that if other countries cheat on their quotas, Saudi Arabia can potentially increase its production so much that the price of oil falls very low. For example, suppose the price fell to £3 per barrel. Saudi Arabia is the only OPEC member making a profit at that price; everyone else loses money.

As a result, Saudi Arabia appears to be in a position to threaten other OPEC members with bankruptcy if they violate their quotas. Unfortunately, the threat doesn't work that well in the real world.

The problem is that Saudi Arabia has limited pumping capacity. Although Saudi Arabia may be able to produce an extra 10 or 20 per cent more oil per day than it normally does, that much of an increase wouldn't be enough to drive the price down to £3 per barrel and bankrupt the other OPEC nations.

As a result, the Saudi Arabian threat isn't nearly strong enough to switch the dominant strategy from cheating on the quota to obeying the quota. And because OPEC has never figured out a way to threaten quota violators effectively, the cartel doesn't work very well.

Regulating Oligopolies

In previous sections, we explain why the Prisoner's Dilemma means that firms in many oligopoly industries have a hard time forming effective cartels. In some industries, however, cartels *are* effective at reducing output and raising prices. Typically, these industries are where one firm is large enough and powerful enough to truly threaten other firms with bankruptcy.

Dealing with dominant firms



In US history, the Standard Oil Company run by John D. Rockefeller during the 19th century dominated an oligopoly industry. The company controlled something like 90 per cent of the oil sold in the United States, and if a competitor didn't do what Rockefeller wanted, he simply bankrupted the other firm by offering oil at a ridiculously low price that the competitor was unable to match.

Rockefeller lost money temporarily while taking this action, but by bankrupting the competitors who disobeyed him, he was able to convince the remaining firms to help him restrain output and make huge profits. Indeed, because Standard Oil exerted so much control, its industry was much more like a monopoly than an oligopoly.

Rockefeller's effectiveness, however, soon brought a governmental response. Standard Oil was broken up into dozens of smaller independent oil companies, none of which was large enough and powerful enough to dominate its industry and enforce collusion as Standard Oil had.

Applying antitrust laws

In the US in the 19th century, cartels were called *trusts* – for example, the Sugar Trust, the Steel Trust, the Railroad Trust and so on. Therefore, laws that broke up monopolies and cartels were called *antitrust laws*. A slightly different approach occurs in the EU where penalties include fines (up to 10 per cent of global turnover for egregious cases) that tend to be levied instead.

Recent cases have included significant fines for cartels in the supply of vitamins by the pharmaceutical industry and undertakings on behaviour by the FA Premier League. The Competition Act and Articles 81 and 82 of the EU Treaty govern this area.

A big problem with antitrust laws is deciding when to regulate oligopolies or break them up to promote competition. The first sign that a cartel may potentially exist is when you see only a few firms in an industry. But, because of the Prisoner's Dilemma, in some cases even a two-firm industry is unable to form an effective cartel. Consequently, prosecutors typically have to do more than just show that only a few firms are operating in an industry.

Typically, concrete proof of collusion needs to be demonstrated. In other words, if one day every firm in an oligopoly decides without co-ordination to cut its output in half and thereby raise prices, that may not be illegal. But if even one email from a manager of one firm to a manager of another firm is found saying that the firms should enter into a cartel, that is illegal and enough for a prosecutor to hang a case on.



A specific and highly effective tactic that's used worldwide is to give immunity from further prosecution to the first firm in a cartel to confess. Given the size of the potential fines, it's very much in your company's interests to blow the whistle as soon as a potential challenge is on the horizon. One case where this happened was in an investigation of British Airways and Virgin Atlantic's use of fuel surcharges: as soon as they became aware of the investigation, Virgin executives promised to make available the incriminating information, in exchange for immunity, resulting in their rival getting the fine (and the break up of the cartel).

Studying a Hybrid: Monopolistic Competition

Monopolistic competition is an interesting form of competition found in certain industries that feature characteristics of monopolies (see Chapter 12) and competitive firms (see Chapter 11).

Benefiting from product differentiation

Like competitive firms operating in free markets, industries featuring monopolistic competition have lots of firms competing against each other. But unlike the situation in competitive free markets, where all the firms sell an

identical product, in monopolistic competition each firm's product is slightly different.



Think about the market for wine. Wine is easily understood as a generic product. However, wine lovers will tell you that they recognise differences between the wines produced by wineries. Although wines are available at different prices reflecting the different processes involved in making the wine, in general the key thing is that the products of any two vineyards are likely to be different from one another in some way.

Economists use the term *product differentiation* to describe the things that make each firm's product a little bit different from its competitors' products. The overall result of these differences is that they slightly decrease the intensity of competition. Your local petrol station, for example, may be able to get away with charging you slightly more per litre than its competitors if it has nice facilities and the next closest competitor is several streets away.

On the other hand, a lot of competitive pressure still exists in the industry. Although your local station may be able to use its unique characteristics to get away with charging you a little more, it can't charge you a lot more – if it tries to do that, you take your business to one of its competitors.

In a similar way, all the restaurants in your neighbourhood have to worry about what the other restaurants are charging, even if the others specialise in completely different cuisines. Although you may be willing to pay 20 per cent more for something exotic, you are unlikely to be willing to pay 90 per cent more. Product differentiation lessens, but doesn't eliminate, price competition.

Facing profit limits

You may think that because monopolistically competitive firms can use their unique characteristics to raise prices, they're guaranteed nice profit margins. After all, in pure competition where firms all sell the same product and have no way of differentiating themselves from their competitors, prices fall so low that firms end up earning zero economic profits (see Chapter 11). If monopolistically competitive firms can raise prices above the competitive price, it seems a no-brainer that they're guaranteed to make profits.

Unfortunately for them, this situation isn't the case.



As pointed out by Cambridge economist Joan Robinson during the 1930s, monopolistically competitive firms still face competition. In particular, they face the prospect that if they're making tidy profits, those profits attract new

entrants to their industry. When the new entrants begin producing, they take business away from the established firms and ruin their previously tidy profits. In fact, new entrants continue to arrive until profits have been driven all the way back to zero.

Dealing with downward-sloping demand

Robinson was able to show how this process works by slightly modifying the monopoly model that we introduce in Chapter 12. To see what she did, look at Figure 13-4, which shows a single monopolistically competitive firm initially making a profit. The figure shows the firm's marginal cost curve, MC , and average total cost curve, ATC , along with its demand curve, D_1 , and the associated marginal revenue curve, MR_1 .

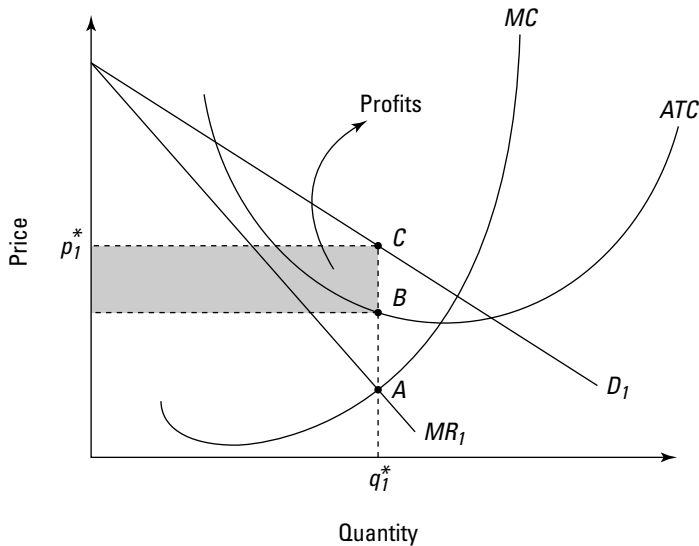


Figure 13-4: A monopolistically competitive firm making a profit equal to the area of the shaded rectangle.

Because of product differentiation, the firm in Figure 13-4 faces the downward-sloping demand curve, D_1 . The firm's demand curve is downward sloping because, like a monopoly, it has some control over its price. Product differentiation means that it can choose whether to set a higher or lower price. At a higher price, the quantity demanded of its product falls because some customers don't think that the unique characteristics of the firm's product are worth the extra money. At a lower price, the quantity demanded increases because the lower price steals customers away from the firm's competitors.

In contrast, competitive firms that sell identical products have no control over the prices they set. As we explain in Chapter 11, because competitive firms sell identical products, the only thing that matters to consumers when choosing among them is who offers the lowest price. The result is that all firms have to sell at the same price, the *market price*, which is determined by where the overall industry supply curve crosses the industry demand curve. The demand curve for an individual competitive firm's product is a horizontal line at the market price (see Chapter 11). This stands in sharp contrast to the downward-sloping demand curve facing the monopolistically competitive firm in Figure 13-4.

An important consequence of the downward-sloping demand curve, D_f , is that the marginal revenue curve, MR_f , associated with demand curve D_f is also downward sloping. Why is this so? The additional, or marginal, revenue that the firm can get from selling an additional unit of output is less than the marginal revenue it gets from selling the previous unit.

As we explain in Chapter 12, declining marginal revenue is a natural consequence of a downward-sloping demand curve. Because the only way to get consumers to buy more of your product is to entice them with a lower price, the marginal revenue you get has to fall with every additional unit you sell.

The monopolistically competitive firm optimises profits by choosing to produce at point A , where the downward-sloping marginal revenue curve, MR_f , crosses the upward-sloping marginal cost curve, MC . Producing the associated quantity, q_f^* , maximises the firm's profit (if making a profit is possible) or minimises its loss. Whether making a profit is possible depends on the position of the firm's demand curve – on how much demand exists for the firm's product.

In Figure 13-4, demand is strong enough that the firm makes a profit. You can see this by comparing the firm's average total cost per unit at output level q_f^* with its selling price per unit at that output level. The average total cost per unit is found by going up vertically from the horizontal axis at output level q_f^* until you hit the ATC curve at point B . The price per unit that the firm can charge at output level q_f^* is found by going up vertically until you hit the demand curve at point C .

Because the vertical distance up to point C exceeds the vertical distance up to point B , you can immediately determine that the firm's selling price per unit exceeds the total cost of production per unit – meaning that the firm must be making a profit on each unit sold. The size of the firm's total profit on all units is the profit per unit times the total number of units sold, so its total profit is equivalent to the area of the shaded rectangle in Figure 13-4. The area of the shaded rectangle is the width of q_f^* units times the height of the profit per unit – the vertical distance between points B and C .

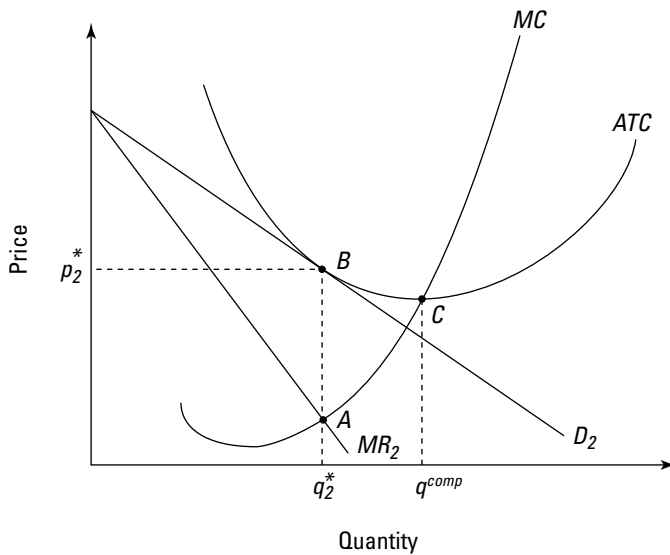
Finding equilibrium: Firm entry and exit

Joan Robinson realised that this profit attracts new entrants to the monopolistically competitive industry. Each new entrant steals some business away from existing firms. Graphically, this results in the demand curve for any existing firm, like that of Figure 13-4, shifting down and to the left. At each possible price that the firm may charge, it sells fewer units than before because new entrants have stolen some of its old business away.

Furthermore, new entrants continue to enter the industry and move demand curves down and to the left until profits are driven all the way down to zero. Only then does the entry of new firms come to a halt.

You can see this sort of equilibrium in Figure 13-5. Here, the demand curve shifts left all the way to D_2 , just tangent to the ATC curve at point B . As the demand curve moves left, so does the marginal revenue curve, which now lies at MR_2 . Consequently, when the firm optimises its production level by producing where the MC curve crosses MR_2 , it now produces at output level q_2^* .

Figure 13-5: A monopolistically competitive firm makes zero profits after entry (or exit) shifts its demand curve until tangent to its ATC curve.



At this output level, profits are zero. You can see this by using the graph to show that the average total cost per unit of producing output level q_2^* is equal to the price per unit that the firm can get selling those units. Go up vertically from the horizontal axis at point q_2^* to point B . Because point B lies both on the demand curve, D_2 , and the average total cost curve, ATC , the vertical distance from the horizontal axis at point q_2^* to point B represents both

the average total cost per unit as well as the price per unit that the firm can charge. They are equal, and so the firm is making zero profits.

If for some reason the firms in a monopolistically competitive industry are making losses, some firms exit the industry. As each of them exits, the remaining firms gain more business, and the demand curves for firms still in the industry shift up and to the right. Exit continues until you reach an equilibrium like that of Figure 13-5, in which all firms are making zero profits.

Producing inefficiently

Importantly, the equilibrium in Figure 13-5 implies that each firm produces less efficiently than firms in a competitive industry. To see this fact, compare the monopolistically competitive firm's output level when the industry is in equilibrium, q^* , with the output level produced by a firm with the same cost curves operating in a fully competitive industry in which all firms sell an identical product. We label this output level as q^{comp} in Figure 13-5.

In Chapter 11, we explain how market forces push competitive firms to produce at q^{comp} , and why this ends up being exactly the level of output at which the *ATC* curve hits its minimum – that is, q^{comp} is the output level at the bottom of the U-shaped *ATC* curve. The socially significant implication of this fact is that competitive firms produce at the lowest possible average total cost per unit. That makes them as efficient as possible in terms of production costs per unit.

In contrast, a monopolistically competitive firm operating in an industry where product differentiation allows it to have some control over the prices it charges ends up producing at a higher average total cost per unit. This situation clearly applies in Figure 13-5, because the vertical distance from the horizontal axis up to point *B* is longer than the vertical distance from the horizontal axis up to point *C*. This fact means that firms in monopolistically competitive industries aren't efficient in the way that firms in competitive industries are.

Some people look at this result and conclude that society may be better off transforming monopolistically competitive industries into competitive industries. But the cost savings may not be worth the loss of product differentiation.

After all, variety is the spice of life. Do you really want every single restaurant to be identical in every way, to serve the same food in the same type of room, under the same lights, with identical furniture? We certainly don't want that. And if the cost of variety is that monopolistically competitive firms produce their output at a higher cost than firms in competitive industries, we're typically willing to endure those higher costs for the sake of having some variety.



Communism, burgers and product differentiation

One of American advertising's most celebrated TV commercials made its debut in 1985, at the height of Soviet communist power. The advert advertises a burger chain called Wendy's, and depicts a communist fashion show. A woman walks down the runway in a drab grey factory worker's uniform, and the announcer shouts out, 'Day wear!' Then she marches down the runway again in the same outfit but this time holding a torch. The announcer shouts out, 'Evening wear!' Next, she marches down the runway again – still in the same uniform – holding an inflatable beach ball: 'Swimwear!'

The commercial made fun of the fact that the central planners who ran communist countries didn't care much about product differentiation.

They typically made only one design of any given product in order to be able to mass produce it at the lowest possible cost. The result was a society in which there was so much sameness that the Wendy's commercial was only a modest exaggeration.

The advert helped to hammer home to US consumers that they should embrace the fact that the food produced by Wendy's was different from that produced by its main rivals, McDonald's and Burger King. Unlike the rigidly planned old Soviet economy, free market capitalism allows for huge amounts of product differentiation. The Wendy's advert emphasised the chain's difference.

But you have to decide for yourself whether you think the high costs of variety are worthwhile – and in what situations. Although the costs of variety may be worth it to you for restaurants, you may have a different feeling about the product differentiation found among petrol stations.

Chapter 14

Property Rights and Wrongs

In This Chapter

- ▶ Defining an ideal market
 - ▶ Seeing how externalities create socially inefficient outcomes
 - ▶ Taking steps to fix problems caused by externalities
 - ▶ Exploiting and exhausting commonly owned resources
-

In Chapter 11, we explain Adam Smith's *invisible hand* – the idea that even though individuals pursue their own interests, if you allow markets to allocate resources, the common good is achieved. Adam Smith was quite aware, however, that to achieve this desirable result, society's property rights must be set up correctly before people start to trade goods and services in markets. In fact, he spent a good deal of his famous book, *The Wealth of Nations*, talking about how governments must properly define property rights if they want markets to yield socially beneficial outcomes.

The gist of the problem is that if property rights aren't set up correctly, a person doesn't fully take into account how his or her actions affect other people. For example, consider two pieces of land. One is privately owned, whereas the other is wilderness land that nobody owns and everyone is free to use as they please. If you want to dump your rubbish on the privately owned land, you have to pay the owner for the right to do so. (In other words, the owner is running a rubbish dump.) But, like everyone else, you can dump your rubbish for free on the wilderness land because no one has the right to stop you.

Naturally, the difference in property rights with respect to the two pieces of land leads people to dump a lot more on the wilderness land because the personal cost to do so is less. But although the cost is less personally, the decision to dump on the wilderness land imposes lots of costs on others. For example, what may have been a very pleasant park becomes a heap of rotting rubbish. Bad property rights lead to bad outcomes.

In this chapter, we talk about positive and negative *externalities* – situations where one person’s behaviour results in benefits or costs to other people, but where the property rights are so badly defined that the costs and benefits aren’t properly accounted for (or alternatively, the cost of a trade between two parties is borne by a third party). Negative externalities result in serious problems such as pollution and global warming. We also show you how most cases of endangered or extinct species are the result of non-existent property rights, and how redefining property rights can save animal species from oblivion.

Allowing Markets to Reach Socially Optimal Outcomes



For markets to achieve socially optimal outcomes, they must take into account all the costs and benefits involved in any activity, regardless of who feels the effects of those costs and benefits. If markets take these factors into account, the demand curve captures all benefits, the supply curve captures all costs and the market equilibrium quantity ensures that only units of output for which benefits exceed costs are produced.

Chapter 11 contains all the details about how supply and demand create socially optimal outcomes, but we want to offer a quick review here. Look at Figure 14-1, which shows a demand curve and a supply curve for ice cream. The market equilibrium quantity is q^* , and the market equilibrium price is P^* .

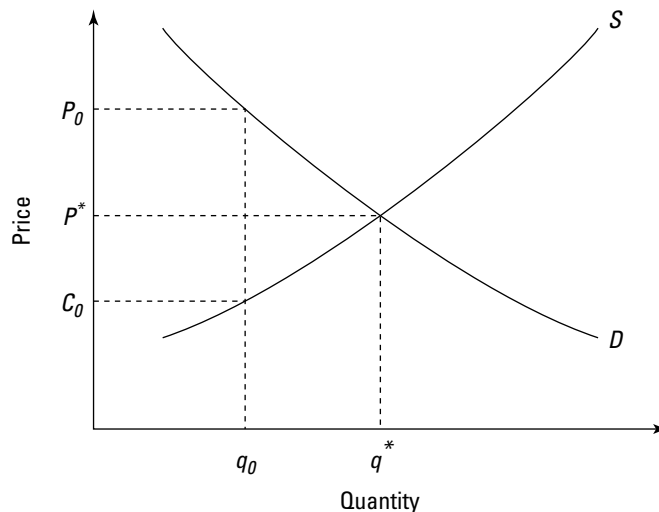


Figure 14-1: With full and complete property rights, the market quantity produced, q^* , is socially optimal.

The important thing to realise is that producing every unit up to and including q^* is socially beneficial. To see the reason why, examine unit q_o . You can see from the demand curve that buyers are willing to pay price P_o for unit q_o , but it costs suppliers only C_o to produce unit q_o .

What does this mean? The overall happiness of society is improved by making unit q_o because people have clearly judged that having the unit is worth the cost of the resources used to make it. Because the demand curve is above the supply curve for all units up to and including q^* , all those units are socially beneficial to produce.



As we explain in Chapter 11, the wonderful thing about markets is that supply and demand just happen to cause the socially optimal level of output, q^* , to be produced. The fact that this result occurs entirely because people are pursuing their own selfish interests is, of course, why markets are so amazing – as though the invisible hand of some kindly deity magically turns the pursuit of individual goals into a socially optimal outcome.

Next, we show you that this desirable result happens only if property rights are *full and complete*, meaning that the demand curve captures all benefits that people are willing to pay for and the supply curve captures all costs associated with production. As you're going to see, if property rights aren't full and complete, markets don't generate socially optimal output levels such as q^* . In such cases, the invisible hand turns out to be *really* invisible – because it doesn't exist!

Examining Externalities: The Costs and Benefits Others Feel from Our Actions

Property rights give owners control over their property. For example, you can paint your car any colour you want. You can modify the engine or the exhaust. You can even install a big booming bass rig, like we may do to try to disguise the fact that, like most economists, we're not actually very cool!

On the other hand, property rights aren't totally unlimited. Society does restrict what you can do with your car. For example, every year cars of a certain age must undergo an MOT for exhaust emissions. Even though you own the car, you can't pump out any more juicy pollution particles than the law permits. You can't drive at 90 miles per hour past a school. And you're breaking the law if you play your 2,000-watt stereo at full volume late at night.

People aren't legally allowed to do these things because we don't live on an island by ourselves. Instead, we live in a community with many other people, and making lots of noise or driving really fast affects their quality of life. The way economists describe this situation is by saying that actions cause *externalities*.

Defining positive and negative externalities

An *externality* is a cost or a benefit that falls not on the person(s) directly involved in an activity, but on others. Externalities can be positive or negative.



- ✓ A *positive externality* is a benefit that falls on a person not directly involved in an activity. Think of a beekeeper. She raises bees to sell the honey, but the bees also happen to fly around pollinating flowers for local farmers, thereby increasing their crop yields and providing them with a positive externality.
- ✓ A *negative externality* is a cost that falls on a person not directly involved in an activity. Think of a steel mill that, as a by-product of producing steel, puts out lots of soot and smoke. The pollution is a negative externality that causes smog and pollutes the air breathed by everyone living near the factory.

Noting the effects of negative externalities

The key thing to understand about negative externalities is that goods and services that impose negative externalities on third parties end up being overproduced. This overproduction happens because negative externalities and the costs that they impose on others aren't taken into account when people make decisions about how much to produce.

Failing to take account of costs imposed on others

In the case of a polluting steel mill, the mill's managers take into account only their private costs of raw materials and running the plant. This situation arises because a poor property rights regime is in place.

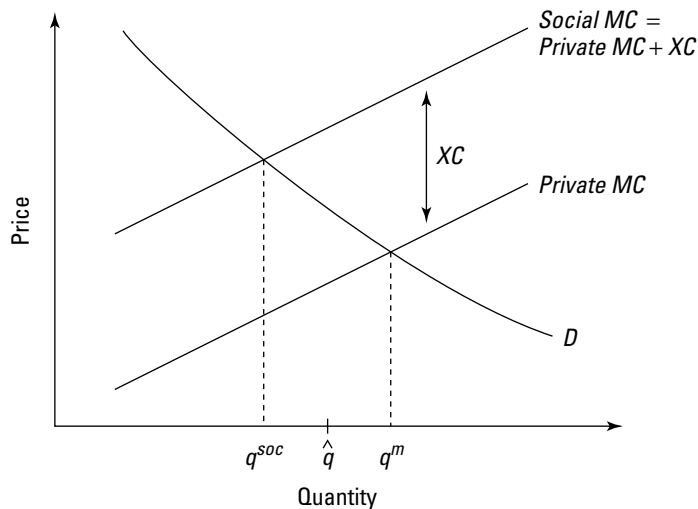
If someone owned the atmosphere, the mill's managers would have to pay for the right to emit pollution. And if the people who had to breathe in the mill's pollution owned the atmosphere, the firm would be forced to pay those people for the right to pollute and be forced to take into account the harm that the pollution causes them.

But because nobody owns the atmosphere and firms don't have to pay to pollute it, no mechanism is in place for making the mill's managers take into account the costs of pollution that fall onto members of the broader community. The result is that the firm overproduces steel.

Why does overproduction happen? In Chapter 10, we explain that a competitive firm's supply curve is equal to its marginal cost curve. Because the mill doesn't take into account the marginal costs that its production of steel imposes on others, its marginal cost curve (its supply curve) is too low and leads to an overproduction of steel.

You can see this situation in Figure 14-2, where we draw two supply curves. The lower one is labelled *Private MC* because the firm's supply curve is its private marginal cost curve, which takes into account only the firm's own costs of producing steel.

Figure 14-2:
The market overproduces a good that has a negative externality.



The higher curve, however, takes into account not only the firm's private costs, but also the external pollution costs, which we label XC (for external costs). This higher curve is called the *Social MC* curve and is useful because it captures all costs associated with producing steel – both the firm's costs of making it and the costs imposed on others as negative externalities.

Overproducing things that impose negative externalities

So why is too much steel produced? The market equilibrium is where the demand curve, D , crosses the *Private MC* curve. That equilibrium results in a quantity q^m of steel being produced, where m stands for market.

On the other hand, the socially optimal amount of steel to produce is q^{soc} , where *soc* stands for social. The point where the *Social MC* curve crosses the demand curve determines the socially optimal quantity. You can tell that q^{soc} is socially optimal because for every unit up to and including q^{soc} , the demand curve is above the *Social MC* curve, meaning that the benefits of producing these units exceed the costs of producing them. These figures apply when taking into account not only private costs, but also the external costs that fall on third parties.

The problem with producing all the units from q^{soc} to q^m is that although the benefits do exceed the firm's private production costs, they don't exceed the total costs when you take into account *XC*, the cost of the negative externality.

For example, look at output level \bar{q} , which lies between q^{soc} and q^m . Go vertically up from \bar{q} to the demand curve and you can see that the market price people are willing to pay for that level of output does exceed the private marginal cost of producing it. (That is, the demand curve is above the *Private MC* curve at output level \bar{q} .) But if you go up even farther, you see that what people are willing to pay for that level of output is actually less than the total social cost of producing that much output. (That is, the *Social MC* curve is higher than the demand curve at output level \bar{q} .)

Output level \bar{q} shouldn't be produced because the total cost of producing it exceeds what anyone is willing to pay for it. That's why the fact that output level q^m is produced in a market economy is unfortunate. Every unit of output produced in excess of output level q^{soc} is a unit for which the total costs exceed the benefits.

Realising that you want positive amounts of negative externalities

A very important thing to realise is that the common reaction to negative externalities – to outlaw them! – is almost never socially optimal. Look back at Figure 14-2 and note that the socially optimal output level q^{soc} is a positive number. That is, producing steel is socially optimal even though some pollution is going to be produced along with it.



To understand the reasoning behind this fact, think about vehicles on the road. Cars pollute the air, and the only way to get rid of this pollution totally is for society to ban all road vehicles. But do you really want to do that?

Although big petrol-guzzling cars certainly produce prodigious amounts of pollution without justifiable benefits, do you really want to get rid of all road

vehicles, including ambulances and fire engines? Not at all, because although these vehicles do emit pollution, the costs imposed on society by the pollution are more than compensated for by their social benefits – the lifesaving activities in which the vehicles are engaged.

The same reasoning holds true for the pollution being produced by the steel factory at output level q^{soc} . The only way to eliminate the pollution completely from the steel factory is to shut it down. But that means removing from society all the benefits that steel can provide, such as earthquake-proof buildings and crash-resistant safety cages in cars.



The goal isn't to eliminate negative externalities. Instead, the goal is to ensure that when *all* costs and *all* benefits are weighed, the benefits from the units of output that are produced outweigh the costs of producing them – including the costs of the negative externalities. In Figure 14-2, for all units of output up to and including q^{soc} , the total benefits are at least as great as the total costs, meaning that society as a whole benefits if these units are produced.

The next thing to consider is how to make sure that only q^{soc} units are produced when, as we show you in the previous section, the market wants to overproduce goods with negative externalities.

Dealing with negative externalities

You can deal with negative externalities in three ways:

- ✓ Pass laws banning or restricting activities that generate negative externalities. For example, most cities now forbid you to dispose of your rubbish by burning it.
- ✓ Pass laws that directly target the negative externality itself (rather than the underlying activity that leads to the externality). For example, steel mills are now required to install *scrubbers* that filter out most of the pollution before it goes into the atmosphere.
- ✓ Impose costs, such as taxes, on people or firms generating negative externalities. For example, the Mayor of London is considering a pollution surcharge on SUVs.

The last of these three solutions is appealing to economists because this solution is most likely to lead to the production of the socially optimal output level.

Look back at Figure 14-2 to see why economists like pollution taxes. Recall that XC is the external cost of the steel mill's pollution on others. If the

government imposes a tax of XC pounds on every unit of steel produced by the firm, the tax raises the firm's cost curve up vertically from *Private MC* to *Social MC*.

Setting the pollution tax at exactly XC pounds causes the firm's marginal cost curve to lie exactly where the *Social MC* curve lies. Because a firm's marginal cost curve is its supply curve, the result is that when demand and supply now interact, the socially optimal output level q^{soc} is produced.

By imposing exactly the right tax on steel, the government can sit back and let the market do the rest. Therefore, this sort of pollution-reducing policy is attractive compared to other potential solutions.

Compare this solution with a system in which firms are ordered to install scrubbers to reduce pollution. In such a system, you need to hire inspectors to monitor factories constantly to make sure that the factories aren't cheating. This sort of system is much more costly to implement than simply imposing a tax on the mill's easily measured steel output and then letting supply and demand set the socially optimal output level.

On the other hand, working out exactly how large the tax XC needs to be to work as an effective deterrent is difficult, and so the pollution-tax solution isn't without problems.

Calculating the consequences of positive externalities

Externalities can be positive as well as negative. The key thing to understand about positive externalities is that goods and services that provide positive externalities to third parties end up being underproduced.

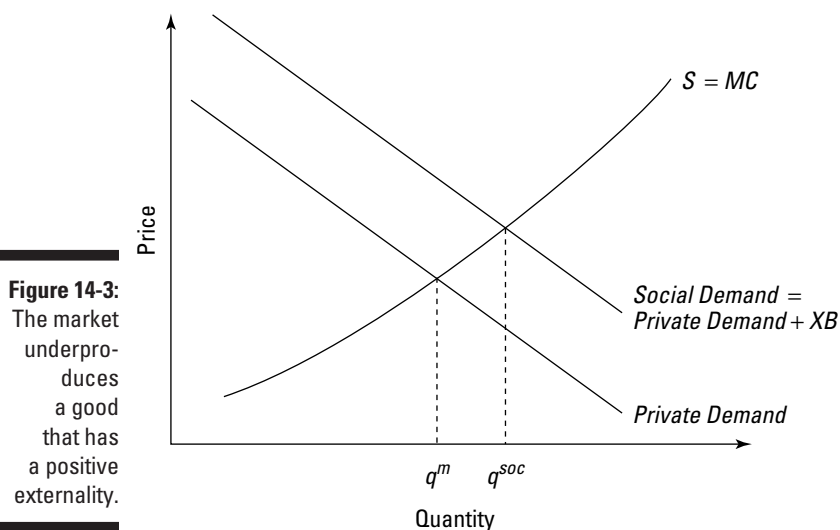
Underproducing things that provide positive externalities



To see why goods that have positive externalities are underproduced, consider a beekeeper named Sally. Sally raises bees so that she can sell the honey and make some money. The people who buy her honey do so because the honey brings them utility when they eat it. But because Sally's bees go around pollinating the flowers of local farmers, these farmers also benefit from her beekeeping activities.

But – and here's the crucial point – the farmers don't pay Sally for the benefits that her bees bring them; the bees just fly in and out of their fields, and no way exists to keep track of them. The result is that Sally is going to raise fewer hives of bees than she would if the farmers were paying her for the benefits that her bees bring them.

You can see how this situation works in Figure 14-3. Sally's supply curve is her marginal cost curve, and we label it $S = MC$. We label the demand for her honey by the customers who pay for it as *Private Demand*. The point where the supply curve and the *Private Demand* curve intersect gives the market equilibrium quantity of honey, q^m .



But this output level doesn't take into account the benefits that bees bring to farmers. Suppose that these benefits have a monetary value of XB , which stands for external benefits. Then the total social demand for Sally's honey is given by the *Social Demand* curve, which is the *Private Demand* curve shifted up vertically by XB pounds to take account of the fact that honey production benefits the farmers as well as Sally's honey-loving customers.

The socially optimal output level, q^{soc} , would be where the *Social Demand* curve crosses Sally's supply curve because, for each unit of output up to and including q^{soc} , the total social benefit is at least as great as Sally's cost of production.

As you can see, the market equilibrium quantity produced, q^m , is less than the socially optimal output level, q^{soc} . In other words, because the market mechanism has no way of taking into account the positive externality, Sally produces less honey than is socially optimal.



Underproduction is typical for goods that generate positive externalities. Because property rights are set up in such a way that the recipients of the positive externalities don't have to pay for them, the producer of the good that generates the positive externalities has no incentive to provide extra

units of output for the benefit of those receiving the externalities. Rather, she produces output only for people who can directly pay her for her product.

Subsidising things that provide positive externalities

Because markets tend to underproduce goods and services that have positive externalities, people have come up with ways to encourage higher levels of production.

The most common way to encourage higher production of goods that generate positive externalities is with a subsidy. In the case of Sally's beekeeping business, the government may actually pay her a honey subsidy of, say, 20 pence per kilogram to encourage her to keep more hives. The result is more bees pollinating more flowers, leading to higher output levels for the farmers. In fact, the government may even tax the farmers to get the money to subsidise Sally's honey: doing so makes the programme pay for itself.

Similarly, governments also often subsidise the planting of trees in and around cities. They must do so because many of the benefits of trees – shade, cooling, cleaner air, less soil erosion and so on – are positive externalities that markets don't take into account. Without the subsidy, fewer trees would be planted than is socially optimal.

Taking in the Tragedy of the Commons

Economists refer to the important economic problem that results from poorly defined property rights, which don't take account of negative externalities, as the *Tragedy of the Commons*. The following sections examine this problem in detail.

Having a cow: Overgrazing on a commonly owned field



To understand the Tragedy of the Commons, think of a farming town in which most of the land is privately owned. However, the town has one large field of *common land* – land on which anyone can graze their cattle. We want to consider the difference between the number of cattle grazed on private land and the number grazed on common land.

In a private field, the owner has an incentive to limit the number of cattle that he puts out to graze. That's because if you put too many beasts in the field, they quickly eat up all the grass and ruin the field for later grazers. Consequently, the owner of a private field puts only a few cattle out to graze.

Doing so reduces his short-run profits (because he restricts the current number of cows), but maximises his long-run profits (because the field stays in good shape, and he can keep grazing cattle well into the future).

On the other hand, think about the incentives that people face when considering the publicly held field on which anyone can graze their cattle. Because the field is commonly owned, you don't have to pay for the right to put a cow out to graze. So everyone is going to want to put some cattle out there because the personal cost of doing so is nothing.

But because everyone is thinking the same thing, the field is soon overrun with cattle and ruined as they eat all the grass and turn it into a big, muddy mess. So, although putting a cow out to graze on a common field incurs no personal cost, a social cost *is* incurred. Each additional cow causes damage to the field – damage that reduces the future productivity of the field.



The difference between what happens to the private field versus the common field is totally the result of the different property rights governing the two types of land. In the case of privately owned fields, farmers have an incentive to weigh the costs as well as the benefits of putting more cattle out to graze. In particular, they take into account how much future profits are going to be reduced if current overgrazing ruins the future usability of the field.

With the commonly held field, however, nobody has a personal incentive to preserve its future usability. In fact, the incentives are actually horribly perverse because if the common field is currently lush with grass, your incentive is to put as many of your own cattle out there as quickly as possible to eat up all the grass before the field is ruined. And because everyone else sees things the same way, a mad rush ensues to put as many cattle out to graze as quickly as possible. The result, of course, is that the field is rapidly ruined – for everyone.

You can view the Tragedy of the Commons as a case of negative externalities. If you see lush grass on the common field, you rush to put out as many of your own cattle as possible without considering the damage that overgrazing causes to the field. The same is true of everyone else. Nobody cares about the negative externality of a ruined field because no individual owns the field and personally suffers when it's ruined.

The important thing to take away from the Tragedy of the Commons is that in the absence of any other regimes or management system, individual incentive to overexploit the common resource tends to form a runaway destruction of the resource. However, please don't take this as meaning that it will inevitably happen in that way. A good example of where it's true is in fisheries, where a Tragedy of the Commons scenario has led to over-fishing of many common species, and has required a system of managed property

rights to correct (we look at this in the next section). However, it actually turns out that many less developed societies have been more effective at using common resources, a point which won Elinor Ostrom the 2009 Nobel Prize in Economics.

You aren't going to be entirely surprised to know that a variant of the Tragedy of the Commons exists that looks at the case of goods that are socially underexploited due to the existence of private property rights. For example, think of the case of the DVD of a TV series that can't be produced because the property rights to the music in the original series are too expensive to license. Because economists are imaginative souls, this variant is known as the Tragedy of the Anticommons. The key thing to take away from this problem is that economists aren't keen on property rights just because they are property rights; economists like them because property rights work when you do a cost-benefit analysis of the situation. In this case, economists would want to test the degree of harm done by the overly strong property rights and recommend modifying rights to mitigate the costs. The principle behind the way we look at the problem, though, remains the same.

Sleeping with the fishes: Extinctions caused by poor property rights

Many environmental problems are caused by Tragedy of the Commons situations in which nobody owns the property rights to a given resource. Notably, most animal extinctions are the result of an absence of property rights.



For example, think of tuna swimming in the open ocean. By international treaty, nobody owns the open ocean. Hence nobody owns the tuna swimming in the open ocean.

On the other hand, if you catch a tuna and pull it up onto your boat, you then have a property right over it and can sell the fish for money. That is, the only way to benefit economically from a tuna is to kill it.

The result is that tuna and many other fish species are hugely over-fished, and many are near extinction, because each fishing boat has the incentive to harvest as many fish as quickly as possible before any other boat can get to them. This situation quickly leads to an extinct species, and the fishing industry is very aware of the problem.

But because of the way that property rights are set up in this case, no single fishing boat can do anything to prevent the calamity. If one fishing boat decides to hold back and take fewer fish with the hope that by doing so the

species is going to survive, someone else just comes in and catches the fish that were spared. The species becomes extinct anyway. As a result, nobody has an incentive to hold back.

Economists look at problems like these and conclude that the only way to stop them is to change the property rights so that people can own a living fish as well as a dead fish. In particular, if you own a school of living tuna, your incentives are very different. You want to preserve the species rather than kill it off, because by preserving it and harvesting fish at a sustainable rate, you benefit not just this year but forever.



Consequently, when economists see a Tragedy of the Commons situation, their first instinct is to change the property rights system governing the resource in question. Instead of commonly held property rights in which each person has an incentive to take as much of the resource as possible before anyone else does, economists suggest private ownership so an incentive exists to preserve the resource.

In the case of over-fishing, one solution has been to give fishers private property rights to an entire fishing ground – that is, to all the fish in an area while the fish are still alive. That gives the new owners the proper incentive to manage the stock on a sustainable basis. Furthermore, because only one person has the right to fish in a given area, a mad rush no longer takes place between competing fishers to harvest as many fish as possible before anyone else can get to them.

For fish species that migrate freely between different areas, a different solution has been developed. In such cases, biologists first determine the maximum number of fish that can be sustainably harvested each year. The government then auctions off fishing permits for exactly that amount of fish.

This method prevents the Tragedy of the Commons by creating a new sort of property right – the fishing permit. The method also has the nice benefit of creating a self-sufficient government programme. The money raised from auctioning off the fishing permits can be used to hire game wardens to prevent unlicensed fishing, as well as for conservation and wildlife management programmes.

A similar technique can be used for carbon emissions (which is already up and running in Europe and in some US states). A 'property right' to a certain amount of emissions is granted by the government, and firms that are more efficient are allowed to sell the surplus emissions that they don't need to less efficient firms. In this way, emissions are not only capped, but also firms are given good incentives to deal with the problem.

Chapter 15

Market Failure: Asymmetric Information and Public Goods

In This Chapter

- ▶ Seeing how secrets can ruin a market
- ▶ Realising how markets are affected when information is incomplete
- ▶ Examining the used car and insurance markets
- ▶ Understanding that markets can't always provide public goods
- ▶ Getting public goods from governments and philanthropists

At this time in history, markets are responsible for providing almost everything that people consume. However, responsible economists would never make the claim that they always function perfectly in every case. In this chapter, we look at what happens when markets – or the market mechanism – fails, a state of affairs economists call – surprise! – *market failure*. In particular, we focus on two of the archetypal reasons for a market failure: asymmetric information and the curious category of things called public goods.

Asymmetric information is a situation in which the buyer knows more than the seller – or the seller knows more than the buyer – about the thing they're bargaining over. The classic example is a high-quality used car: the owner who's trying to sell the vehicle knows all about the car's high quality and reliability, but the potential buyer can only take the owner's word for it.

Because the potential buyer has no reason to trust the seller's assertions that the car is really good, he assumes the worst and offers a low price just in case the car turns out to be a lemon (the technical term for an old banger). But because the owner knows that the car is of high quality, she rejects the low offer and the car goes unsold – all because no cheap and easy way exists to prove the car's high quality to the potential buyer. This isn't just a textbook case either, as asymmetric information lies at the heart of the Credit Crunch, as we will see later in the chapter.

Public goods cause market failure in a different way, because the very nature of a public good counts against a private seller being able to charge users. The classic example is a lighthouse. When the building is up and running, it benefits all nearby ships, regardless of whether they pay for the service.

Therefore, each and every ship owner tries to avoid paying for the service in the hope that somebody else may pay for it. But with everybody thinking the same way and leaving somebody else to pay, the lighthouse soon goes bankrupt and society is denied a valuable service.

In the remainder of the chapter, we discuss these two causes of market failure in more detail, show you how pervasive they are and describe some of the clever solutions that people have come up with to remedy them. So don't expect any asymmetric information here (sorry!).

Facing Up to Asymmetric Information

Many situations exist in real life in which buyers and sellers don't share the same information. Depending on the situation, the buyer or the seller may be better informed.



For example, when it comes to selling used cars, sellers are much more knowledgeable about the true quality of the vehicles than the buyers. On the other hand, when it comes to home insurance, the buyers of the insurance policies are much better informed because they know all about their homes, and the security methods they employ to safeguard them.

Regardless of which party is better informed, economists refer to these situations as cases of *asymmetric information* because one side has more information than the other. (Bear in mind that information can be incomplete, but *symmetric* when each party is equally in the dark about the other.)

Realising that asymmetric information limits trade



Asymmetric information is very important in the real world because it limits exchange in a market. If you know that the other person is better informed, you're afraid that he may use his information to take advantage of you. Similarly, if you cannot expect their honesty in the deal, you will be less likely to make one.

In the case of used cars, buyers are afraid that sellers who know their cars are bad may keep that fact to themselves and try to negotiate high prices – as

if their cars were good. In the case of insurance, insurance companies are afraid that people who know they are high insurance risks may pretend to be low insurance risks so that they can get lower rates.

Depending on how bad the asymmetric information gap is, markets may even collapse completely. That is, if you have huge worries that the seller of the used car may be exaggerating the value of the vehicle she's trying to sell you, you probably aren't going to buy. That sounds like a reasonable thing to do, but it prevents the sale of good cars because everybody's worried about bad cars. Similarly, if insurance companies can't figure out a way to tell the good insurance risks from the bad insurance risks, they may charge high rates to everybody as though everyone is a high risk. And that, typically, causes the low-risk people not to buy insurance because they know they're being overcharged.



So keep in mind that asymmetric information can lead to *market failure* – and in the worst cases completely collapse markets as bad goods drive out good. In these cases, no market can exist because people have been scared off by the fact that other market participants are better informed and have the means to use that against them.

Souring on the lemons problem: The used car market and the credit crunch

Berkeley economist George Akerlof received the Nobel Prize in Economics in 2001 for a famous paper he wrote called 'The Market for Lemons'. The paper is all about asymmetric information and market failure, and it was especially memorable because Professor Akerlof used the market for used cars as his primary example.



The used car market was interesting to Akerlof because it suffered from an interesting form of market failure: almost all the used cars for sale are lousy (lemons). What Akerlof correctly explained was that poor-quality vehicles, or lemons, dominate the market because asymmetric information drives away almost all sellers who want to part with high-quality used cars.

To make the logic behind the result clear, imagine that only three kinds of used cars are available for sale: good, okay and bad. They all look the same on the outside and even test-drive pretty much the same, but they have major differences in terms of how much longer they're going to last before the engine gives out. Because of the difference in engine quality and how long the cars are likely to last, the good cars are worth, say, £15,000, the okay cars are worth £10,000 and the bad ones are worth only £5,000.

The problem that leads to market failure is the asymmetric information that exists between buyers and sellers. In particular, although each seller knows how good their own car's engine is, the buyers have no way of knowing.

Buyers can, of course, ask sellers to be truthful about the quality of their cars, and no doubt many sellers – probably most – would tell the truth. But no way exists to *know* if they're telling the truth. Consequently, when a particular seller tells you that their car is good, you're still going to be nervous about being cheated.

As we're about to show you, this very reasonable fear causes nearly all good and okay cars to be withdrawn from the market. The result is a used car market that's dominated by bad cars; as Akerlof put it, the used car market ends up becoming 'a market for lemons'.

Seeing how quality used cars are driven from the market



Imagine that you want to buy a used car, but you don't want to overpay for it. You know that only three types of cars exist: good, okay and bad. In addition, you happen to know that one-third of all used cars on the road is good, one-third is okay and one-third is bad. How much are you willing to pay for a used car?

Well, given the fact that good cars are worth £15,000, okay cars are worth £10,000 and bad cars are worth £5,000, and also given the fact that you don't know which cars are which, imagine that you're willing to pay no more than £10,000.

Why £10,000? Because that's what the okay used car – the car of average quality – is worth.

Because sellers have no way of proving to you how good their cars are, a sensible thing to do when presented with a used car is to assume that it's of average quality and, therefore, worth £10,000. So you offer £10,000. And so do all the other buyers presented with used cars because they, like you, can't tell the quality of used cars apart.

Now look at how different sellers react to the £10,000 offer depending on the true quality of their vehicles:

- ✔ If a seller knows his car is bad and worth only £5,000, he happily accepts your offer.
- ✔ If a seller knows his car is okay, he also accepts because you're offering what the car's actually worth.
- ✔ If a seller has a good car, he doesn't accept unless he's in some sort of dire circumstance. He knows the car is worth £15,000, and so he refuses to accept your £10,000 offer unless he's really desperate to raise cash in a hurry (perhaps to pay off some gambling debts to a guy named Machete Bob).

The result is that nearly all the good cars on the market are withdrawn, leaving only bad and okay cars. Now consider how that situation changes what buyers are willing to offer.

If all the good cars are withdrawn from the market, a 50/50 chance now exists that a car is okay or bad. In such a case, how much are you going to offer to pay if you're a buyer? Well, with a 50/50 chance of the car being worth £10,000 or £5,000, you probably offer the average of these two values: £7,500. But when you do, the market becomes even more dysfunctional.

After all, how are the sellers of okay cars going to react to being offered £7,500? They're going to reject the offer and withdraw their vehicles from the market too.

The sad result is that with the good cars and then the okay cars withdrawn from the market, the only cars left are the bad ones, the lemons. Because of the asymmetric information problem, the used car market ends up being a market for lemons.

Buyers are aware of this situation, and so they offer only £5,000 for any car on the market. And because only bad cars are offered, sellers accept the £5,000. So although bad cars end up being priced correctly in the used car market, no market exists for good or even okay used cars.

That's a problem because people – both buyers and sellers – want to trade (and would be much happier trading) good and okay cars. But unless some solution can be found to the asymmetric information problem, they're all left out in the cold.



If you've got to this point, you'll have followed the logic as it applies to one type of real-world goods market. But this example underplays the effect asymmetric information might have on the world at large – after all, there is a readily available substitute for second-hand cars, called new cars. But what if we depended on a market continuing to supply something even as asymmetric information is driving providers out of the market? And what if that were something as necessary to the continued functioning of an economy as credit?

Following Lehman Brothers' bankruptcy filing in September 2008, creditors began to wonder whether Lehman was the only poor debtor in town. (As we now know all too well, they were not.) Soon, the financial sector began to withdraw credit from the market as sums were done and exposures calculated. Lenders could not be confident that borrowers were not lemons, and so they began to withdraw their custom from the market, exactly as Akerlof's model predicted. This withdrawal of credit from the market is what we now call the *credit crunch*. Whilst it wasn't the ultimate cause of the banking crisis, it had a big hand in creating the panic that ensued.

Making lemonade: Solutions to the lemons problem

The fundamental issue driving the lemons problem is that the sellers of good and okay cars have no way of convincing buyers that their cars are as good as the sellers know them to be. The whole problem would be resolved if some way was found to convince buyers that a good car is, in fact, a good car and an okay car is, in fact, an okay car.

In the next three sections, we discuss ways to achieve this goal. These methods don't work perfectly, but because they offer some reassurance to buyers, buyers are willing to offer enough to get sellers to part with higher quality cars.

Offering a warranty

One way a seller can convince a buyer that she's really got a good car is to offer the buyer a warranty. A warranty is convincing because only the seller of a good car is willing to offer a warranty. The seller knows that her good car isn't going to break down after the sale, meaning that she's never going to have to pay for any repairs.

On the other hand, the seller of a bad car would never offer a warranty because he knows that his car is likely to break down and that he's going to have to pay for the repairs.

Consequently, if someone is willing to offer you a warranty, she almost certainly has a good car. That's why you see so many used car dealerships offering warranties on the vehicles they sell. If they don't offer warranties, the lemons problem quickly takes over, and prices fall so low that only bad cars can be bought and sold on the used car market.

Building a reputation

Another way to solve the lemons problem is to reassure buyers by setting up a market in such a way that sellers can build a reputation for honesty and fair dealing. For this reason, most good used cars are sold through used car dealers rather than directly between individuals.

Compare a used car dealership with an individual selling her used car online. Who has more of an incentive to tell you the truth about car quality?

The used car dealer makes his living selling used cars, so if he overcharges one customer by pretending a bad car is good, he soon gets in trouble. When that car starts to break down, the buyer becomes angry and tells all his friends that the dealer cheated him. And that loss of reputation cuts into the dealer's future sales. In fact, he's quickly going bankrupt if he develops a reputation for lying.

In contrast, an individual selling her used car doesn't have to worry about developing a reputation for lying. Her main source of income isn't selling cars. If she cheats you and you get mad and tell all your friends, it doesn't affect her much because she's not in the business of selling used cars.

The result is that she has much more of an incentive to lie than the used car dealer who has to worry about his reputation. As a result, most good used cars are sold through used car dealers. (But even at used car dealers, people still need some reassurance, which is why most used car dealers also offer warranties.)

As the great Marxist (Groucho, that is) said: 'The key to success in life is honesty and fair dealing. If you can fake those, you've got it made.'

Getting an expert opinion

Because the heart of the lemons problem is asymmetric information, another method of resolving the problem is for sceptical buyers to hire an expert who can give them the information they need to distinguish good, okay and bad cars. Many car buyers employ this strategy when they have doubts about a seller's honesty.

For a fee that is small relative to the total purchase price, a buyer can hire a disinterested third-party expert – for example, a professional mechanic – to inspect the vehicle and make a list of repairs that are most likely needed in the near future. In this way, the buyer can get a better picture of the car's quality and what a fair price may be.

However, this method isn't fully able to resolve the asymmetric information problem because the expert probably can't discover *everything* that may be wrong with the car. To the extent that this situation is true, buyers may still be suspicious, and some potential for market failure may remain. For this reason, you often see buyer-initiated inspections used in conjunction with other methods of resolving asymmetric information, such as warranties and sales by dealers who have a reputation to protect.

Issuing insurance when you can't tell individuals apart

An insurance company faces an asymmetric information problem of its own: the people buying insurance know more than the company about the risks they face.



Consider car insurance. Who needs it more: good drivers who hardly ever get into accidents, or bad drivers who get into lots of accidents? Now, clearly, even good drivers want insurance because they're sometimes involved in accidents for which they're not to blame. But bad drivers want insurance even more to help pay for all the accidents they know they're going to cause because of their poor driving. Economists call this problem *adverse selection*.

An asymmetric information problem faces the insurance companies because although individual drivers know whether they're good or bad, the insurance companies can't easily tell them apart. If they were able to tell them apart, insurance companies would simply charge the good drivers a low rate for insurance and the bad drivers a high rate.

But because they can't tell the good and bad drivers apart, the insurance companies run a serious risk of going bankrupt. To see why, imagine that insurance companies offered the same low rate to everyone, as though they were *all* good drivers. This strategy soon leads to bankruptcy because the insurance companies aren't collecting enough in premiums to pay off all the damage caused by the bad drivers.

To avoid bankruptcy, the insurance companies may go to the other extreme, charging everyone as though they were bad drivers. But then the good drivers stop buying insurance because for them it's overpriced. The result is that only bad drivers sign up for insurance.

This result is very poor for society because you want everyone to be able to buy insurance at a rate that fairly reflects his or her driving ability. Good drivers should be able to get insurance at a fair rate. And because good drivers make up most of the drivers in the real world, insurance companies lose out on lots of potential profits, unless they can figure out a way to separate the good drivers from the bad drivers.

Grouping individuals to help tell them apart

Insurance companies have come up with a paradoxical way of dealing with the fact that they can't tell whether an individual is a good or bad driver. Instead of focusing on the individual, they look for clues about the individual based on the groups to which he or she belongs. Doing so often gives the insurance companies a pretty good idea about whether the individual is a good or bad driver.

For example, males aged under 25 are known to get into many more accidents than females under 25. So if two people walk into an insurance company and one of them is a 23-year-old male and the other is a 22-year-old female, chances are that the male is a much worse driver than the female. Consequently, you charge the male a higher insurance rate.

This situation has the nice result of making sure that everybody can get insurance at what is *likely* to be a fair price given the fact that, on average, males under 25 get into many more accidents than females under 25.

In reality, this nice result isn't the compelling reason behind the decision of insurance companies to look at what groups their customers belong to and discover as much as possible about them. These companies really have no choice; competition *forces* them to do so.



Why is this true? Consider two insurance companies, only one of which uses group-membership information to help set rates. The company that doesn't use group information has to set very high rates because of the fear that all its customers may be bad drivers. Doing so drives away all the good drivers who don't want to pay bad-driver rates for their insurance.

But the company that uses group information can offer multiple rates, such as high ones to young men and low ones to young women. Doing so allows the company to capture the business of many good drivers who don't want to deal with the first insurance company that sets only one high rate for everyone. The result is that insurance companies are always looking for ways to try to estimate an individual's unknown risk profile based on the well-known risk profiles of the groups to which he or she belongs.

This process can lead to some rather unfair conclusions. The oddest result is that good-driving young males end up paying higher rates than bad-driving young females because the only thing insurance companies have to go on is gender.

But such a system is still better than the even more unfair alternative in which *all* good drivers have to pay bad-driver rates, which is what would happen if insurance companies were banned from using group-membership information to try to distinguish their customers. The closer insurance companies can get to fully distinguishing good and bad drivers using group-membership information, the fairer are the rates.

Keep in mind that the drivers for whom insurance companies have the greatest need to use group-membership information for are new drivers. Because insurance companies don't have any accident or violation records for new drivers, a pressing need exists to try to separate the good from the bad drivers using group-membership information. As drivers get more experience, the insurance companies can get increasingly accurate accident and violation information that distinguishes the good from the bad.

Avoiding adverse selection

Using the groups to which a person belongs to estimate his or her individual insurance risk goes only part of the way to resolving the asymmetric information problem that exists between insurance companies and their customers.

Obviously, any group still features a lot of individual variation. For example, even if young women are, on average, better drivers than young men, some young women are bad drivers. If an insurance company sets a premium for young women on the basis of how often young women *on average* get into accidents, insurance is more attractive to young women who are really bad drivers than to young women who are really good drivers.

As a result, bad-driving young women are more prone to sign up for insurance than good-driving young women. This tendency is known as adverse selection because the bad, or adverse, insurance risks seem to self-select into buying insurance policies. The result is a customer pool that contains a disproportionately high number of bad drivers.



Adverse selection is a difficult problem because it feeds on itself. The insurance company has to raise rates to take account of the fact that bad drivers are more likely to sign up than good drivers. But when it raises rates, the problem just gets worse because the higher rates make insurance even less attractive to good drivers, meaning that the pool of applicants is going to be even more disproportionately dominated by bad drivers.

One solution to adverse selection is for an insurance company to offer a large group of people one rate – on the condition that nobody can opt out. For example, at one American university, a health insurance company offers the university one low rate for every employee on the condition that every employee must be enrolled. By enrolling everyone, there's no chance that the less healthy are going to dominate the insurance pool because all the healthy have declined to be enrolled.

Mitigating moral hazard

The other big problem facing insurance companies is called *moral hazard*. Moral hazard arises because buying insurance tends to change people's behaviour. For example, if you don't have car insurance, you're likely to drive much more slowly, knowing that you have to use your own money to pay for any damage you cause. But because you do have insurance, you may drive faster and more recklessly knowing that if something goes wrong, the insurance company is going to be stuck with the bill. Similarly, because you have contents insurance, you may be more prone to leaving your door unlocked. (Please know that when we say 'you' in examples like this one, we don't actually mean *you*. We're quite sure that *you're* above moral hazard even if your friends and neighbours aren't.)

Car insurance companies deal with moral hazard by offering discounts in exchange for high excesses. For example, if you get into an accident, the £1,000 excess that you've chosen means that you have to pay the first £1,000 of any bills resulting from the accident.

EXAMPLE



Group discrimination, individual identification

The idea of grouping individuals to help sort them extends beyond insurance. For example, companies want hard-working employees but can't tell when you walk in for an interview if you are, in fact, hard-working. So they try to estimate the chances that you are hard-working by seeing what groups you belong to.

For example, nearly all straight-A students are hard-working. Therefore, if you're a straight-A student, a company is going to be much more likely to hire you. You may actually be lazy, but

by seeing what group you belong to, the firm improves its odds that you're not.

The practice of using information about the groups to which an individual belongs to try to figure out personal characteristics is referred to as *statistical discrimination*. Although this practice typically improves economic outcomes, you have to decide for yourself whether – and in what cases – you think the practice is fair or unfair.

The excess serves as a strong inducement for you not to give into moral hazard and drive recklessly. And because the insurance company knows that your high excess gets rid of most of your moral hazard problem, they're willing to offer you insurance at a lower rate than if you opt for only a £100 excess.

Excesses are a clever way of reducing moral hazard problems and helping to make insurance more affordable for responsible drivers.

Providing Public Goods

Public goods are things that private firms can't profitably produce because no way exists to exclude non-payers from using them. The inability of private firms to produce public goods profitably derives from the fact that public goods have two very special characteristics: they are non-rival and non-excludable.

EXAMPLE



- ✓ *Non-rival* means that one person using the good doesn't diminish another person's ability to enjoy the good. Think of a fireworks display, a statue in a park or a television show broadcast over the airwaves. Your consumption doesn't in any way diminish that of others. This situation stands in stark contrast to most goods, where if you consume more, less remains for others. (Think of cookies.)

✓ *Non-excludable* means that preventing non-payers from consuming a good or service is difficult. For example, when you produce a fireworks display, everyone in the vicinity gets to see it for free, no matter how much you may want to charge them for it. A more serious example is an army: when an army is in place to provide national defence, it provides national defence for everybody, including those who don't want to contribute to the cost of maintaining it.



The non-rival and non-excludable characteristics of public goods make it very hard for private firms to make any money producing them. Think about trying to get people to buy tickets to an outdoor fireworks display. Because people know that they're going to be able to see it for free, they don't buy tickets. Because they don't buy tickets, there's no way to raise the money needed to put on the display.

This sort of chicken-and-egg problem is frustrating, because though people don't want to pay for something they can get for free, they actually do like firework displays – meaning that they're fundamentally willing to pay something to see them. The problem is working out how to get them to pay.

Taxing to provide public goods

The most common solution to the problem of how to provide public goods is for governments to step in and use tax money to pay for them. In the case of fireworks, because nearly everybody likes fireworks, there's no problem getting enough political support for spending tax money on displays. And after the taxpayers have funded them, everyone can enjoy the fireworks.

The government has historically provided national defence because it too is very much a non-excludable, non-rival public good. For example, because protection from foreign invaders is non-excludable, a temptation exists not to help pay for it because you know that if someone else does pay, you get to enjoy safety from foreign invaders for free. And because national defence is non-rival, you know that the safety you enjoy is of just as high a quality as the safety everyone else enjoys. This fact lessens your incentive to pay too. As a result, governments force people to share the expense for national defence by levying taxes.

Taxes and a good portion of government spending are often derided as wasteful (and indeed they may be), but keep in mind that public goods are only likely to be provided through universal provision via taxes. Although nobody likes taxes, unless you're a strict libertarian, you probably wouldn't want to trade a reduced tax burden for no public parks, no national army, no public fireworks displays, no public roads, no public sewers and so on.

Without the government's ability to force people to pay for these things, people probably wouldn't have them – at least not in the quantity and variety that people currently enjoy.

Enlisting philanthropy to provide public goods



Although government taxation pays for most public goods, some goods are paid for privately. For example, in the city of Dundee you'll find a number of public parks originally donated to the city by the Jute Barons like the Cox brothers, who gained their wealth in the thriving textiles industry of the 19th century. The public goods (the parks) were therefore provided at private expense.

In ancient Greece and Rome, public philanthropy went even further, with rich aristocrats building roads, aqueducts and temples for public use. In some cases, rich men even paid for entire armies to be sent out to defend the country in times of war.

So please don't think that governments are absolutely necessary to provide public goods. They aren't. But they are a much more reliable way to provide public goods because you don't have to rely upon the philanthropic largess of the rich, who are under no obligation to spend their wealth on public rather than private goods.

Along the same lines, don't make the common mistake of thinking that public goods are called public because the government rather than the private sector typically provides them. Economists call them public goods because private firms can't profitably produce them, not because they have to be produced by the government. Private philanthropy can produce public goods without any help at all from the government.

Providing a public good by selling a related private good



Broadcast television is a public good. After a TV programme's signal is sent out over the airways, that programme is non-rival: the fact that you're watching the programme doesn't reduce anyone else's ability to tune in. The programmes can also be seen as practically non-excludable: even if you can stop someone from tuning in without an access card or without paying a licence fee, broadcasters have to worry about people receiving the programmes by

other (generally illegal) means. So given that TV programming is very much a public good, why are lots of TV programmes produced by privately owned and operated TV stations?

The answer is that the US broadcast industry figured out that although TV itself is a public good, the adverts that accompany TV programmes are very much private goods for which they can charge a lot of money. If a car maker, beer maker or the publisher of a revolutionary new economics book with a yellow and black cover wants its ad to be shown to the millions of viewers who tune in for free to the public good known as TV, that company has to pay for commercial air time.

The trick behind TV is that the privately sold good called advertising pays for the freely provided public good called television. To a more limited extent, newspapers work the same way. Although they raise some money from subscriber fees or the newsstand price, a huge chunk of their revenue comes from the advertising they sell. This model is now commonplace as a business method for most old and new media and even some forms of software.

Of course, advertiser funding isn't the whole of the story. In most countries, a balance exists between state sponsorship and private provision. In the UK, the BBC, funded from a licence fee, is a solution to the non-excludability condition. Over time, the UK slowly moved towards a mixture of funding solutions for different channels, including subscription, advertising, pay per view and combinations of the above. However, the principle that everybody pays the licence fee and this payment supports 'Auntie' has remained constant.

Ranking new technology as a public good

We live in an age of rapidly rising living standards. Why is this the case? Because institutions are fostering the creation of new and better technologies that allow us to produce more goods and services from the same old resources, or to produce entirely new goods and services that were previously impossible to produce.

Technological progress is a public good. Therefore, society has had to come up with ways to make sure that technological progress happens given the fact that private individuals and firms have little incentive to invent new technology.

To understand how new technologies are public goods, consider the invention of the moveable-type printing press by Gutenberg in 1435. Before Gutenberg, books were copied by hand. But after he invented the printing press, printing new copies became much cheaper.



Furthermore, think about how simple the technology really is. The printing press is basically just a big version of the rubber stamps that little kids like to play with. The invention was immediately understandable to anyone and everyone who heard about it, which meant that they were able to make their own printing presses as soon as they heard about the invention.

So how does the invention of the printing press satisfy the characteristics of a public good?

- ✔ The invention is non-rival because building and using a printing press doesn't in any way lessen anyone else's ability to build and use a printing press.
- ✔ The invention is basically non-excludable because the cost of communicating the new idea to another person is so low – just a short conversation does the trick.

The result is that unless society creates some sort of institution to reward the creation of new ideas, very little profit incentive exists to going into the invention business. In fact, what happened to Gutenberg was that everyone copied his idea and didn't pay him for it. So unless you can come up with a way to reward the creation of new inventions financially, you're unlikely to get many inventions.

Patenting to turn public goods into private goods

The solution has been the creation of patents. By giving inventors the exclusive right to market and sell their inventions for 20 years, patents provide a financial incentive to get people to invest the time and energy necessary to come up with new technologies that benefit everyone. The fact that the industrial revolution took off only after government-enforced patents became widely available in western Europe in the 18th century is not coincidental.

Subsidising research into technologies that can't be patented

But even today, not every new innovation can be patented, because you can patent only something you invent, not something you discover. For example, if you think up a chemical that's never existed before and then synthesise it, you can patent it. But if you merely discover an existing chemical that's been floating around the sea or lying in the soil, you can't patent it.

This situation is a big problem for things such as cancer research because many potential cures are chemicals derived from plants and animals, chemicals that have existed in nature for eons. These chemicals have huge potential benefits, but because they can't be patented, nobody has a strong financial incentive to try to discover them.

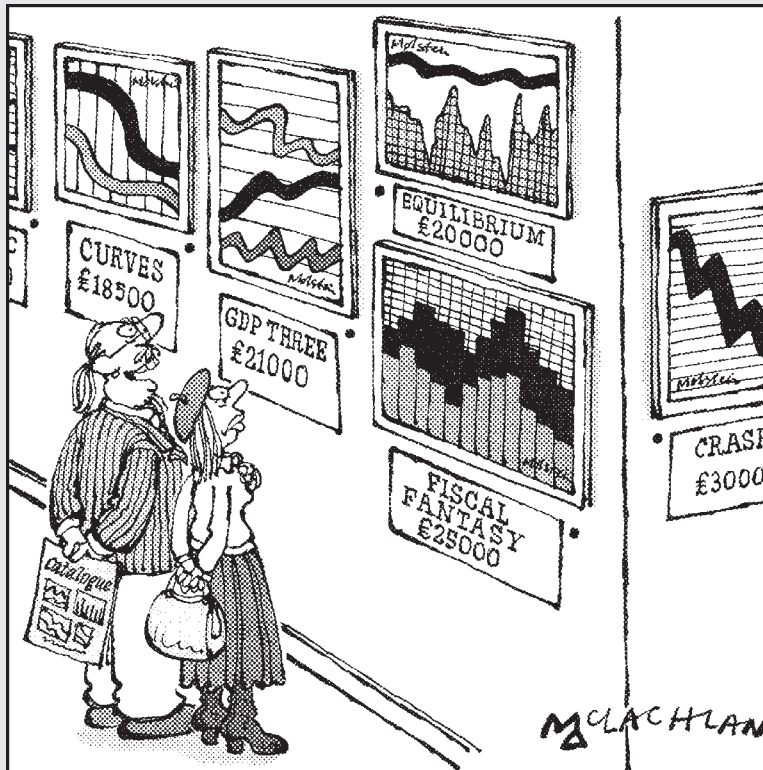
As a result, the government and many private philanthropic groups fund research into areas of science where the public goods problem would otherwise limit research.

This solution is very important to an economist because providing public goods is an economic problem that markets and the invisible hand can't fix. Other types of market failure, like asymmetric information, have pretty decent private sector solutions (as we discuss in the section 'Making lemonade: Solutions to the lemons problem', earlier in the chapter).

But unless a society can come up with good ways of providing public goods, that society is permanently deprived of their benefits. For public goods like fireworks displays, this hardly matters. But for technological innovations like curing cancer, the situation is literally a matter of life and death.

Part IV

The Part of Tens



'He was a lousy economist but his charts caught the eye of a famous art critic & dealer.'

In this part . . .

The chapters in this part offer some fun, fast reading. Chapter 16 covers the lives and ideas of some great economists. Chapter 17 debunks ten very common but incorrect pieces of economic thinking – the kind of stuff you hear pundits and politicians refer to all the time. Chapter 18 features ten correct and truly great economic ideas that need to guide your thinking about public policy and how best to run an economy.

Chapter 16

Ten (Or so) Famous Economists

In This Chapter

- ▶ Realising that no economist works in isolation
 - ▶ Singling out some amazing economists
-

In this short chapter, we give you brief descriptions of the ideas put forth by twelve of the very best and most influential economists (ten just wasn't enough). Each person radically changed the way that economics conceptualises the world or radically changed the way that politicians and government officials formulate public policy.

But don't for a second think that these men did it all on their own (and, yes, they are all men – history works like that sometimes). As with any science, a single person's breakthrough in economics is built on the foundation of hundreds of contributions made by scores of researchers.

In other words, a whole lot more than ten – or even twelve – great economists have existed. With any luck, this book has sparked your interest to discover more about economics, so you can come to know the stellar ideas of the many great economists who didn't happen to make this list.

Adam Smith

Adam Smith (1723–1790) was a professor of Moral Philosophy who developed the idea that as long as robust competition constrains firms, their self-interested profit-seeking inadvertently causes them to act in ways that are socially optimal – as though they're guided by an *invisible hand* to do the right thing. He also analysed the way that new technology and organisation combined to create new and more efficient ways of doing things, leading to technical and economic progress.

But Smith wasn't naive. He believed that people who run businesses prefer to collude rather than compete whenever possible, and that governments have a very important economic role to play in fostering the robust competition needed for the invisible hand to work its magic. He also believed that governments must provide many essential public goods, such as national defence, that aren't readily produced by the private sector.

David Ricardo

David Ricardo (1772–1823) discovered the concept of comparative advantage and argued (correctly) that international trade is a win-win situation for the countries involved. Comparative advantage destroyed the intellectual respectability of *mercantilism*, the mistaken theory behind colonialism that viewed trade as being one-sided and consequently argued that trade should be set up to benefit the mother country at the expense of its colony.

In addition, Ricardo correctly analysed the economic phenomenon of diminishing returns, which explains why costs tend to increase as you increase production levels. He was also a strong early proponent of the quantity theory of money, the idea that increasing the money supply increases prices.

Karl Marx

Karl Marx (1818–1883) was the foremost economist among 19th-century socialists. Few of his major economic theories are now believed to be true, but because proponents of his Marxist ideas came to power in dozens of countries during the 20th century, he is surely one of the most influential economists who ever lived. (Marx gets the most space here not because he's the most important economist on this list, but because we have to take the time to explain his ideas before discrediting them. The ideas of the other economists on this list are already explained in detail in other places in this book. Also, when reading about Marx, bear in mind that a fair question to ask is: just how Marxist were 20th-century Marxist governments?)



Marx's most important intellectual contribution is his idea that capitalism is a historically unique form of social and productive organisation. In his book *Capital*, he analysed capitalism as a brand-new form of social and economic organisation based on capital accumulation and factory production. He called the owners of the factories 'capitalists' and argued that they would be forced to exploit the workers who laboured in their factories.

In particular, he believed that the only capitalists who would survive and whose businesses would grow were those who paid workers the minimum salaries necessary for the workers to survive. Thus, even as productivity and output rose rapidly, workers would endure permanent, grinding poverty out of which they'd never be able to rise except by means of a violent overthrow of the capitalists – an overthrow in which the workers would gain control over the factories.

Marx argued that this violent overthrow would be facilitated by what he saw as an inevitable tendency toward concentration and monopoly. When only one monopoly firm in each industry existed, the workers would find it much easier to revolt and take over the system.

With a century and a quarter of hindsight, we know that Marx was wrong in his economic thinking. In particular, workers' wages *do* rise over time – in fact, they rise on average as fast as technological innovation increases productivity levels. That's because capitalists compete over the limited supply of workers, and wages get bid up as quickly as productivity improvements allow one capitalist to bid higher wages to steal workers away from other capitalists.

In addition, competition does *not* lead to each industry being dominated by a single monopoly firm, and even if that were inevitably true, governments would still have a strong interest in preventing that outcome. Instead, competition remains robust in most industries, and consequently delivers all the benefits of Adam Smith's invisible hand.

Alfred Marshall

Alfred Marshall (1842–1924) invented the supply-and-demand method for analysing markets. Applying mathematics to economic theory, he clearly differentiated between *shifts* of demand and supply curves and movements *along* demand and supply curves. In doing so, he cleared up 2,000 years of faulty reasoning. He also made the revolutionary prediction that the market price would be where the demand and supply curves cross.

Marshall then went one step farther and realised that by comparing points along demand and supply curves with the market price, you can quantify the benefits that consumers and producers derive from market transactions. These benefits are, respectively, consumer surplus and producer surplus, and their sum is the total economic surplus.

This method of quantifying the benefits of production and consumption is still used today and forms the basis of *welfare economics*, which studies the

costs and benefits of economic activities. This method also just happens to illustrate in a very simple graph the reasoning behind Adam Smith's invisible hand. The free market equilibrium, where demand and supply cross, is exactly what a benevolent social planner would choose to do if she were trying to maximise social welfare by maximising total economic surplus. In other words, a free market does indeed act 'as if moved by an invisible hand' to promote the common good.

John Maynard Keynes

John Maynard Keynes (1883–1946) invented modern macroeconomics and the idea of using government-provided economic stimuli to overcome recessions. Much of the rest of 20th-century macroeconomics was a series of responses to his seminal ideas.

His most famous ideas were developed in response to the long agony of the Great Depression of the 1930s. He first asserted that the Great Depression was the result of a collapse in the expenditures being made on goods and services. He then asserted that monetary policy had been ineffective in combating the decline in expenditures. And he finally concluded, given his dismay about monetary policy, that fiscal policy was the only remaining source of salvation. In particular, Keynes believed that the best way to increase expenditures in such dire circumstances was for the government to spend heavily to pay for programmes that would buy up lots of goods and services in order to get the economy moving again.

Keynes's policy prescriptions were adopted during the Great Depression in many countries and later formed the bedrock of the post-war welfare state. And although many of his specific ideas about the cause of the Great Depression and the best policies for dealing with recessions are no longer embraced, his underlying idea that governments are responsible for taming the business cycle remains very much with us today.

Kenneth Arrow and Gerard Debreu

Kenneth Arrow (b. 1921) and Gerard Debreu (1921–2004) mathematically proved that Adam Smith's idea of the invisible hand was, in fact, correct. Not only do competitive firms provide society with the utility-maximising combination of goods and services, but also they do so efficiently, at minimum cost. Since this proof came in the 1950s, it served to disprove the assertions of totalitarians and communists that centrally planned economies were more productive or more efficient than market economies.

Milton Friedman

Milton Friedman (1912–2006) convinced economists that the quantity theory of money is, in fact, true: sustained inflations are the result of sustained increases in the money supply (printing too much money). This insight put limits on using monetary policy to stimulate the economy.

Friedman also argued that the Great Depression was chiefly a monetary disaster and that its severity was the result of a gruesomely tight money supply that kept real interest rates much too high. This diagnosis of the cause of the Great Depression is now the standard explanation, meaning that the intellectual ammunition for Keynes's solution to recessions – large increases in government spending – has lost much of the sway that it once had. Friedman's diagnosis has also led economists to conclude that monetary policy is more important than fiscal policy for regulating the economy and preventing recessions.

Paul Samuelson

Paul Samuelson (b. 1915) has made many contributions to economics. Perhaps the most important was crystallising the idea that all economic behaviour can be thought of as consumers and firms maximising utility or profits subject to a set of constraints. This idea of *constrained maximisation* has become the dominant paradigm that governs how economists conceive of economic behaviour.

Samuelson also developed a judicious blending of Keynesian and classical ideas about the proper use of government intervention in the economy. Keynes argued for large government interventions to mitigate recessions. Classical economists such as Smith and Ricardo argued for minimal government interventions, fearing that government interventions tend to make things worse.

Samuelson's *neoclassical synthesis* states that during recessions the government needs to be willing to make large interventions in the economy to get it moving again, but when the economy is operating at full potential, the proper role of government is to provide public goods and take care of externalities. Many economists embrace this view of the government's place in the economy.

Robert Solow

Robert Solow (b. 1924) has made huge contributions to the understanding of economic growth and rising living standards. In addition to developing innovative models of how economies grow over time, he also showed that the dominant long-run force propelling economic growth is technological innovation.



Before Solow, the economic profession believed that increases in output were the result of increases in inputs. In particular, increases in output were solely the result of using more workers or more capital (such as bigger factories). What Solow demonstrated was that *at most* 50 per cent of the long-run growth of living standards can be explained by increases in labour and capital. The rest has to be the result of technological innovation.

This insight created a huge paradigm shift among economists that has resulted in the systematic study of technological innovation and the ways in which it can be improved by government policies such as patents, or by investment in human capital. Solow's insight also opens up the refreshing possibility that technological innovation can allow us to enjoy higher living standards without having to constantly increase our use of the earth's resources.

Gary Becker

Gary Becker (b. 1930) has been hugely influential because he pushed economics into areas that were previously immune to economic thinking.

His first major contribution was to argue that free markets tend to work *for* equality and *against* racial and gender discrimination. The logic is that firms that refuse to hire the best-qualified workers because of their race or gender put themselves at a competitive disadvantage relative to non-biased firms. Becker backed up this insight by showing that industries that are more competitive do, in fact, employ more minorities and women.

Another significant contribution that Becker made was to model families as economic units in which family members tend to act on the basis of cost-benefit analyses. For example, as societies became richer and paid employment became more plentiful (and better paying), Becker predicted that more women would choose to work rather than stay at home. He provided an economic explanation for a huge change in the labour force that otherwise would have been explained only in terms of sociological considerations (such as changing gender roles).

Similarly, he was the first to model criminal behaviour in terms of how criminals view the potential costs and benefits of committing any given crime. If the expected benefits exceed the expected costs, the criminal is most likely going to attempt the crime. This theory of criminal behaviour is radically different from previous explanations, and it led Becker to propose the very influential idea that the best way to deter crime is to raise the costs relative to the benefits.

Robert Lucas

Robert Lucas (b. 1937) showed that people are sophisticated planners who constantly modify their optimal strategies in response to changes in government policy. If you assume that people change their behaviour only very slowly in response to policy changes, you're likely to overestimate the results of those changes.



In particular, monetary policy loses most of its effectiveness if people rationally plan for policy changes. Suppose the government announces that in three months it's going to double the money supply in an attempt to stimulate increased purchases of goods and services. If shop owners keep prices the same despite the fact that more money is on the way, the economy is stimulated because people can buy a lot more stuff with all that new money.

But if, instead, shop owners rationally react to the announcement, they're going to raise their prices in anticipation of all the new money that's going to be spent in their stores. By doing so, they greatly reduce the amount by which sales of goods and services increase when people begin to spend all the new money.

In particular, if the shop owners double their prices in anticipation of the doubling of the money supply, the policy change isn't going to result in any increase in the amount of goods and services sold. With prices twice as high, having twice as much money allows customers to buy only exactly as much as they did before.

Lucas's idea came to be known as *rational expectations*, and brought with it a new humility about the extent to which government policy – monetary policy in particular – can influence the world.

Chapter 17

Ten Seductive Economic Fallacies

In This Chapter

- ▶ Avoiding logical fallacies that catch out intelligent people
 - ▶ Steering clear of bad economic reasoning
-

In this short chapter, we outline the most attractive and compelling incorrect ideas in economics. Some are logical fallacies. A few are myopic opinions that don't take into account the big picture. And others are poorly thought-out examples of economic reasoning. All these ideas are to be avoided.

The Lump of Labour Fallacy

The argument that a fixed amount of work exists, which you can divide up among as many people as you want, is often presented as a cure for unemployment. The idea goes that if you convert from a 40-hour work week to a 20-hour work week, firms have to hire twice as many workers. In 2000, for example, France reduced its work week to only 35 hours in the hope that firms would hire more workers and cure France's persistent unemployment problem.

The idea didn't work; such policies have never worked. One problem is that hiring workers involves many fixed costs, including training costs and health insurance. So two 20-hour-per-week workers cost more to employ than one 40-hour-per-week worker. What's more, two 20-hour-per-week workers don't produce any more output than one 40-hour-per-week worker.

So if laws were passed that forced firms to move from a 40-hour work week to a 20-hour work week, firms wouldn't double the size of their workforces. They'd hire fewer than twice as many workers because costs would go up, which is one reason why, in 2005, France relaxed its experiment with the 35-hour work week.

In addition, even if cutting the work week in half actually did double the number of workers used, it would only hide the overall unemployment problem by spreading it around. If 100 per cent of workers are working half-time, they are all 50 per cent underemployed. That situation is not a significant improvement over having 50 per cent of the population employed full-time and 50 per cent unemployed.

What you really want is a situation in which every worker who wants a full-time job is able to get one. Shortening the work week doesn't achieve this goal.

The World Is Facing an Overpopulation Problem

Various versions of the overpopulation myth have been floating around since the late 18th century when Thomas Malthus first asserted the idea. He argued that living standards can't permanently rise because higher living standards cause people to breed faster. He believed that population growth would outpace our ability to grow more food, so we would be doomed to return to subsistence levels of nutrition and living standards.

Even when Malthus first published this idea, lots of evidence indicated that it was bunk. For generations, living standards had been rising while birth rates had been falling. And because that trend has continued up to the present day, we're not going to breed our way to subsistence.



Indeed, many nations now face an *under*population problem. In most developed countries, birth rates have fallen below the replacement rate necessary to keep the population stable. As a result, their populations are soon going to start shrinking dramatically. And because birth rates are falling quickly all over the world, the United Nations expects the total human population to max out at around 9 billion people in 2070.

A related problem is that rapidly falling birth rates are wreaking havoc on government-sponsored retirement systems because too few young workers exist to pay all the taxes needed to fund retirees' pensions. In desperation, some countries are going so far as to pay cash bounties to mothers for each new child they give birth to. (Italy, which has had one of the lowest birth rates as well as one of the most generous pension systems, is particularly worried.)

Although many countries with relatively high birth rates do have poverty and malnutrition problems, many economists are just as willing to ascribe poverty to poor government policies.

Again, this argument isn't to deny that the world faces many environmental problems over the coming century. However, as people become more educated and aware of their environmental impact, overpopulation is becoming less of a concern.

The Fallacy of Confusing Sequence with Causation

Post hoc ergo propter hoc is a Latin phrase that translates roughly as, 'Because you see one thing precede another, you think that it causes the other.' That is, if A happens before B, you assume that A causes B.

Such a deduction is false because A and B often don't have any relationship. For example, if it rains in the morning and you get a headache in the afternoon, that doesn't mean that the rain caused your headache.

Politicians try to pull this logical fallacy all the time when discussing the economy. For example, suppose that politician A gets elected, and a few months later a recession hits. The two may have nothing to do with each other, but you can be sure that during the next election, an opponent of politician A claims that the recession was the result of politician A's policies. The only proof offered is that one event happened before the other. In these cases, it's worth looking for good reasons to support the argument!

Protectionism Is the Best Solution to Foreign Competition

To be sure, you can find both good and bad arguments in favour of protectionism. The good arguments include the need for strategic security for certain goods. For example, most developed countries place a degree of protection on their defence industries. However, arguments in favour of trade barriers and taxes on imports, on the grounds that these policies benefit citizens and prevent jobs from being exported, tend not to be good arguments. The problem is that their arguments consider only the benefits of protectionism without also considering the costs.

Trade barriers and taxes on imports *do* protect the specific jobs that they're intended to protect. However, other jobs are often sacrificed in the process.



For example, raising tariffs on foreign coal protects the jobs of domestic miners. But such a policy results in higher energy costs all over the economy. Domestic manufacturers have to pay higher energy costs than if they had access to the cheaper foreign coal, and so they have to raise the prices of the goods they produce. As a result, demand for these goods decreases, and the manufacturers don't need as many employees.

Another problem with protectionism is that citizens are consumers as well as producers. For example, if the government prevents the importation of lower cost, higher quality foreign cars, it preserves jobs in the domestic car industry. But costs for domestic consumers rise as a result.

Protecting an unproductive industry that faces foreign competition only allows it to keep using resources that would be better used by more vibrant industries. Workers who would otherwise move to jobs in innovative, highly productive new industries instead get stuck in an industry so unproductive that it can survive only by having the government rig the economy in its favour.

Granted, the move from a dying industry to an innovative new industry can be rough for an individual worker. But instead of protecting unproductive industries to avoid the need for change, the government can help domestic workers more efficiently by providing retraining programmes for employees. (In the case of older workers who have only a few years of employment left, early retirement programmes may be more viable than retraining.)

The Fallacy of Composition

Assuming that what's good for one person to do is good for everyone to do all at once is another common fallacy. For example, if you're at a sold-out sporting event and want to get a better view, standing is a good idea – but only if you're the only one who stands up. If everyone else also stands up, everyone's view is just as bad as when everyone was sitting down (but now everyone's legs are getting tired). Consequently, what was good for you to do alone is actually bad for everyone to do at the same time.



The fallacy of composition is false because some things in life have to do with relative position. For example, if you start out as the lowest-paid employee at your firm but then get a 50 per cent rise while nobody else gets a rise, your relative position within the firm improves. However, if everyone gets a 50 per

cent rise at the same time, you're still the lowest-paid person at the firm. If what matters to you is your relative standing within the firm, getting the same rise as everyone else doesn't make you any happier. On the other hand, if you are more interested in where you stand relative to people who work at other firms, getting a 50 per cent rise is good even if everyone else at your firm gets it too! (Funnily enough, the most recent research suggests that when we think about what makes us better off, we're overwhelmingly likely to be motivated by doing better than our peers!)

If It's Worth Doing, Do It 100 Per Cent

We all value safety. But was a famous US politician really being sensible when he said that we should spend whatever money may be necessary to make flying on commercial airlines 'as safe as possible'?

Economists would say, 'No!' The problem is that making commercial airline travel 'as safe as possible' would mean making it prohibitively expensive. Although safety is a good thing, achieving complete safety is not a worthy goal if doing so makes flying so expensive that only the extremely wealthy can afford it.

The politician failed to apply *marginalism* – the idea that the best way to approach a problem is to compare marginal benefits with marginal costs. Applying marginalism to airline safety, you realise that making flying 'as safe as possible' is wasteful.

The first few airline safety innovations (such as seatbelts and radar) are sensible to undertake because the extra, or marginal, benefit that each brings is greater than the extra, or marginal, cost required to pay for it. But after the first few safety innovations are implemented, successive innovations become more costly and less effective. At some point, additional innovations bring only small marginal increases in safety while running up high marginal costs.

We're not suggesting that investments in safety are not worthwhile. Although increased airport security does hurt airlines, it's something we can accept because the risks associated with someone getting through with bad intentions are so high that we are willing to pay the cost. However, even here there's a point where we'd draw the line and say no more. After all, if no one flies, terrorists cannot attack air flights, but the costs to the global economy would be enormous.

Free Markets Are Dangerously Unstable

Free markets are volatile because supply and demand often change very quickly, causing rapid changes in equilibrium prices and quantities (which we discuss in Chapter 8). Rapid change isn't a problem, however. The responsiveness of markets is actually one of their great benefits. Unlike a government bureaucracy that can never react quickly to anything, markets can adjust to huge changes in world events in only minutes.

The new equilibrium prices and quantities ensure that resources are allocated to their best uses and that society suffers from neither shortages nor gluts. So don't call markets unstable. Call them *responsive*.

Low Foreign Wages Mean That Rich Countries Can't Compete

You often hear that UK firms can't compete with firms based in developing countries because of vast differences in hourly wages. To see the problem with this thinking, compare a factory in Uttoxeter with a factory in Phnom Penh.



Say the UK factory pays its workers £20 per hour while the factory in Cambodia pays £4 per hour. People mistakenly jump to the conclusion that because the foreign factory's labour costs are so much lower, it can easily undersell the UK factory. But this argument fails to take into account two things:

- ✓ What actually matters is labour costs per *unit*, not labour costs per *hour*.
- ✓ Differences in productivity typically mean that labour costs per unit are often nearly identical despite huge differences in labour costs per hour.

To see what we mean, compare how productive the two factories are. Because the UK factory uses much more advanced technology, one worker in one hour can produce 20 units of output. The UK worker gets paid £20 per hour, so the labour cost *per unit of output* is £1. The factory in Cambodia is much less productive; a worker there produces only 4 units in one hour. Given the foreign wage of £4 per hour, the labour cost per unit of output in Cambodia is also £1.

Obviously, Cambodia's lower hourly wage rate *per hour* doesn't translate into lower labour costs *per unit* – meaning that Cambodia is unable to undersell its UK competitor.

People who focus exclusively on labour costs omit some of the most important parts of the analysis. The cost of production also depends on factors such as the availability of capital, machinery and resources, or the regulations imposed by the host government. Economists disagree on the exact impact of these factors, although they do agree that labour on its own is insufficient to account for differences in the cost of producing items in different countries.

Keep in mind that governments can seriously screw up what would otherwise be a near equality of labour costs per unit by fixing artificially low exchange rates. For example, if at an exchange rate of 8 Chinese yuan to 1 US dollar labour costs per unit are equal, the Chinese government can make labour costs per unit look artificially low to US consumers if it fixes its currency at, for example, 16 yuan to 1 dollar. In such situations, the inability of US workers to compete with Chinese workers is due to the currency manipulation, not to the lower wage rate per hour found in China. For this reason, among others, the US has been attempting to persuade China to revalue the yuan. Another example of this tendency is Italy (again!) who, until prevented from doing so by entering the euro, used competitive devaluations of the lira to boost exports (leading to the need for a hefty number of zeroes on even the smallest notes). At the time of entry to the euro, £1 was worth around 2,500 lira.

Tax Rates Don't Affect Work Effort

Some politicians argue for raising income taxes as though the only effect of doing so is to raise more money. But experience has demonstrated over and over again that beyond a certain point, people respond to higher taxes by working less. And that reduction in labour denies society all the benefits that would have come from the extra work. (Also, because people work less, the increased tax rate doesn't bring in nearly as much revenue as expected.)

So if you see a politician arguing for an increase in income taxes, look into the details to make sure that the disincentive effects of the tax hike don't cause more mischief than the benefits that are going to be derived from spending the money raised by the tax increase. (Note that this section isn't an argument against tax, because we've already established that taxes can have very positive effects. The situation is, though, a reason for doing cost-benefit analysis on tax changes to make sure that the result isn't unnecessary economic damage.)

Forgetting That Policies Have Unintended Consequences Too

When evaluating a policy, people tend to concentrate on how the policy is going to fix some particular problem while ignoring or downplaying the possible other effects. Economists often refer to this situation as *The Law of Unintended Consequences*.



Suppose that you impose a tariff on imported steel in order to protect the jobs of domestic steelworkers. If you impose a high enough tariff, their jobs are indeed protected from competition by foreign steel companies. But an unintended consequence is that the jobs of some autoworkers are lost to foreign competition. Why? The tariff that protects steelworkers raises the price of the steel that domestic car makers need to build their cars. As a result, domestic car manufacturers have to raise the prices of their cars, making them relatively less attractive when compared to foreign cars. Raising prices tends to reduce domestic car sales, meaning that some domestic workers in the car factory lose their jobs.

Unintended consequences are far too common. Be aware of them whenever a politician tries to persuade you to see things his or her way. Chances are that the politician is mentioning only the good results of a certain policy; he or she may not even have thought about the not-so-good side effects.

Chapter 18

Ten Economic Ideas to Hold Dear

In This Chapter

- ▶ Understanding basic economic principles
 - ▶ Arming yourself against the economic follies of politicians
-

In this chapter, we list ten economic ideas that all informed people need to understand and be ready to use to evaluate the policy proposals made by politicians. Some of these ideas aren't necessarily true in all situations, but because they are usually correct, be wary if someone wants you to believe that they don't apply to a particular situation. Make that person convince you, because the chances are that he or she is wrong.

Society Is Better Off When People Pursue Their Own Interests

This concept is basically Adam Smith's famous invisible hand. If all economic interactions in a society are voluntary on the parts of all parties involved, the only transactions that are going to take place are those in which all parties feel they are being made better off.



If you trade your gold for another person's bread, you're likely to do so because you value his bread more than your gold. You trade because trading makes you better off. Meanwhile, you can be sure that the other person values your gold more than his bread. So trading makes him better off too. You both pursue your self interests, and you're both made better off.

This concept of what motivates people doesn't mean that charitable acts are bad for society. Instead, it means that even philanthropy is generated by self-interest. People give because they enjoy helping others. By doing so, both they and the people they help are made better off.

Free Markets Require Regulation

Economists firmly believe that voluntary transactions in free markets tend to work toward the common good. But they also believe that nearly every participant in the marketplace would love to rig the system in his or her own favour. Adam Smith, in particular, was quick to point this out and argue that for markets to work and serve the common good, the government has to fight monopolies, collusion and any other attempts to prevent a properly functioning market in which firms vigorously compete against each other to give consumers what they want at the lowest possible price.

Economic Growth Depends on Innovation

At any given moment, a fixed amount of wealth exists that can be divided equally among all people, like slicing a pie into equal pieces and giving each person one equal slice. But if living standards are to keep rising, you need a bigger pie to split up. In the short run, you can get a bigger pie by working harder or using up resources faster. But the only way to have sustained growth is to invent more efficient technologies that allow people to produce ever more from the limited supply of labour and physical resources.

Freedom and Democracy Make Us Richer

Very good moral and ethical reasons exist for favouring freedom and democracy. But a more bottom line reason is that, in general, because freedom and democracy promote the free development and exchange of ideas, free societies have more innovation and, consequently, faster economic growth.

Education Raises Living Standards

Educated people not only produce more as workers – and hence get paid higher salaries – but also, more importantly, they produce innovative new technologies. Sustained economic growth and higher living standards are only possible if you educate your citizens well. Of course, other good reasons exist for getting an education, including the ability to appreciate high art and literature. But even if all you care about is living in a country that has rising living standards, you should work hard to promote education in the sciences and engineering, sectors where revolutionary technologies are created. (Notice that we don't say that lots of people should become economists. All economists can do for growth is to urge others to become engineers!)

Protecting Intellectual Property Rights Promotes Innovation

People need incentives to encourage them to take risks. One of the biggest risks you can take is to leave a secure job in order to start a new business or work at developing a great new idea. Intellectual property rights, when deployed effectively, give you a bargaining chip to help ensure that the rewards are going to go to you and your associates rather than competitors. Without this assurance, fewer people would be willing to take the personal risks necessary to provide society with innovative new technologies and products.

Weak Property Rights Cause Many Environmental Problems

People always have to do some polluting. After all, even if you don't want gas-guzzling SUVs running around causing lots of pollution, you probably still want ambulances and fire engines to operate despite the fact that they too pollute the environment. The difference is that the overall benefit to society outweighs the cost of the pollution in the case of the emergency vehicles but not in the case of the SUVs.

Seen in this light, society's goal isn't to ban pollution completely, but to make sure that the benefit exceeds the cost for whatever pollution is generated. As we discuss in Chapter 14, strong property rights are key to ensuring that people weigh the complete costs and benefits of pollution. Property rights force people to take into account not only their personal costs of generating pollution, but also the costs that their actions impose on others.



Because nobody owns the atmosphere, you don't have to pay anyone for the right to pollute. Polluting the air is, in fact, free – which leads to much too much polluting.

In contrast, we can't just throw rubbish anywhere because somebody owns every bit of land in the world. If we want to throw rubbish on someone's land, we have to pay that person for permission or risk huge fines (or even prison) for dumping rubbish without permission. Also, because we have to pay rubbish collection fees to throw out our rubbish, we're discouraged from generating wasteful amounts of it.

All environmental problems tend to stem from poorly defined or non-existent property rights that allow polluters to ignore the costs that they impose on others. Therefore, economists favour the creation and enforcement of property rights systems that force people to take all costs into account.

International Trade Is a Good Thing

Opening your country to international trade means opening your country to new ideas and new innovations. Competition from foreign competitors causes local businesses to innovate to match the best offerings of companies from around the world.



Quite simply, throughout history, the richest and most dynamic societies have been the ones open to international trade. Countries that close themselves off from international trade grow stagnant and are quickly left behind. Of course, what economists have in mind when they think of the benefits of international trade is *free trade*, where companies compete across borders to provide people with the best goods and services at the lowest prices. Economists strongly condemn the many government subsidies and trade restrictions that impede free trade and that try to rig the game in one country's favour.

Free Enterprise Has a Hard Time Providing Public Goods

Private firms can provide goods and services only if they can at least break even doing so. To break even (or make a profit), whatever a firm is selling has to be *excludable*, by which we mean that only those paying for the good or service receive it.

As we explain in Chapter 15, some goods and services are non-excludable. For example, a lighthouse provides warning services to all ships in the vicinity regardless of whether they pay the lighthouse keeper. Because every ship knows that it can get the service without having to pay for it, the private lighthouse quickly goes bankrupt because only a few ships are fair-minded enough to pay for the service.

Goods and services that are non-excludable are called *public goods* because they're essentially open to the public and can't be kept private.

Because private firms can't make a profit producing public goods, you typically need governments to provide them. Unlike private firms, governments can force people to pay for public goods. They do this by levying taxes and using the tax revenues to pay for public goods, such as the army, the police force, lighthouses, public fireworks displays, basic scientific research and so on.



Economists view the existence of public goods as one of the most important justifications for government intervention in the economy. Although private philanthropy can also provide some public goods, many public goods are so expensive that they can be provided only if the government uses its power of taxation to fund them. Consequently, public goods are typically publicly provided.

Preventing Inflation Is Easy (ish)

Governments can cause high rates of inflation by increasing the money supply too rapidly. A growing economy always has a growing demand for money because with more stuff to buy, you need more money with which to buy it. If you want to keep the overall level of prices constant, the correct response is to increase the money supply at the same rate that demand is increasing. If the supply of money increases faster than the demand for money, the value of money falls, creating inflation. (In other words, it takes more money to buy the same amount of stuff as before, meaning that prices go up.)

One recent development has been to make central banks like the Bank of England independent. This move gets around the temptation to use monetary policy as an instrument of growth by making it the central banker's job to ensure that prices stay stable no matter what governments decide.

Appendix

Glossary



aggregate demand: The total demand for goods and services in an economy.

aggregate supply: The total supply of goods and services in an economy.

allocative efficiency: A term describing a situation where the limited resources of an economy are allocated to the production of the goods and services that consumers most greatly desire to consume.

antitrust laws: Laws that regulate *monopolies* and *cartels*. Used interchangeably with competition law (more common in the UK).

asymmetric information: Situations in which the buyer or the seller knows more about the quality of the good that they're negotiating over than the other party does.

capital: Machines, factories, and infrastructure used to produce output.

cartel: A group of firms that colludes and acts as a single co-ordinated whole to restrict output and drive up prices.

command economy: An economy in which the government directs all economic activity.

comparative advantage: The argument developed by David Ricardo that countries should specialise in producing those goods and services that it can produce most cheaply within the economy and import those that are more expensive to produce.

competitive market: A market in which many sellers compete against each other to attract customers. Each seller has an incentive to sell at the lowest price possible to attract customers, so prices tend to be driven so low that the sellers can just barely make a profit.

Consumer Price Index (CPI): National Statistics *market basket* used to measure changes in the prices of goods and services bought by a typical household; complements and in some cases replaces the Retail Price Index (RPI).

consumer surplus: The benefit consumers get when they can buy something for less than the maximum amount that they are willing to pay for it.

deadweight loss: The amount by which *total surplus* is reduced whenever output is less than the *socially optimal output level*.

deflation: When the overall level of prices in the economy is falling.

demand: The whole range of quantities that a person or group of people with a given income and preferences demands at various possible prices.

demand curve: A line on a graph that represents how much of a good or service buyers are going to consume at various prices.

depreciation: A decrease in the economy's stock of *capital* caused by wear and tear or obsolescence.

diminishing marginal utility: A situation where each additional, or marginal, unit of a good or service that you consume brings less *utility* than the previous unit.

diminishing returns: A situation where each additional amount of a resource used in a production process brings forth successively smaller amounts of output.

economic costs: Total costs, including money spent on production and *opportunity costs*.

economic model: A mathematical simplification of reality that allows you to focus on what's really important by ignoring lots of irrelevant details.

economic profits: Any monies collected by a firm above and beyond what is required to keep an entrepreneur owner interested in continuing in business; also known as rents.

economics: The study of how people allocate scarce resources among alternative uses.

externality: A cost or benefit that falls not on the person(s) directly involved in an activity, but on others. Externalities can be positive or negative.

factors of production: Inputs (resources) used to create goods and services, including land, labour, capital, and entrepreneurship.

financial markets: Markets where people trade the property rights to assets (like real estate or stocks) or where savers lend money to borrowers.

fiscal policy: A government's policy on taxes and spending.

fixed costs: Costs that have to be paid even if a firm isn't producing anything.

full-employment output (Y^*): How much output is produced in the economy when full employment exists in the labour market.

gross domestic product (GDP): The value of all goods and services produced in the economy in a given period of time, usually a quarter or a year.

human capital: The knowledge and skills that people use to help them produce output.

hyperinflation: When the inflation rate exceeds 20 or 30 per cent per month.

increasing returns: A situation where each additional amount of a resource used in a production process brings forth successively larger amounts of output.

inflation: When the overall level of prices in the economy is rising.

inflation rate: A measure of how the overall level of prices in the economy changes over time. If the inflation rate is positive, prices are rising. If the inflation rate is negative, prices are falling.

interest rate: The price you have to pay to borrow money.

investment: Any increase in the economy's stock of *capital*.

invisible hand: Adam Smith's famous idea that when constrained by competition, each firm's greed causes it to act in a socially optimal way, as if guided to do the right thing by an invisible hand.

Law of Demand: The fact that, for most goods and services, price and quantity demanded have an inverse relationship.

long-run shutdown condition: A situation where a firm's total revenues exceed its *variable costs* but are less than its total costs. The firm continues to operate until its *fixed cost* contracts expire (in the long run).

macroeconomics: The study of the economy as a whole, concentrating on economy-wide factors such as interest rates, inflation, and unemployment. Macroeconomics also encompasses the study of economic growth and how governments use monetary and fiscal policy to try to moderate the harm caused by recessions.

marginal cost: How much total costs increase when you produce one more unit of output.

marginal utility: The change in total *utility* that results from consuming the next unit of a good or service. Marginal utility can be positive or negative.

market basket: A bundle of goods and services selected to measure inflation. Economists define a market basket, such as the *Consumer Price Index*, and then track how much money it takes to buy this basket from one period to the next.

market economy: An economy in which almost all economic activity happens in markets, with little or no interference by the government; often referred to as a *laissez-faire* ('leave alone') economic system.

market failures: Situations where markets deliver socially non-optimal outcomes. Two common causes of market failure are *asymmetric information* and *public goods*.

market production: Term that economists use to capture what happens when one individual offers to make or sell something to another individual at a price agreeable to both.

markets: Places where buyers and sellers come together to trade money for a good or service.

microeconomics: The part of economics that studies individual people and individual businesses. For people, microeconomics studies how they behave when faced with decisions about where to spend their money or how to invest their savings. For businesses, it studies how profit-maximising firms behave individually, as well as when competing against each other in markets.

monetary policy: Using changes in the money supply to change interest rates in order to stimulate or slow down economic activity.

monopolistic competition: A situation in which many firms with slightly different products compete. Production costs are above what may be achieved by perfectly competitive firms, but society benefits from the product differentiation.

monopoly: A firm with no competitors in its industry. A monopoly firm produces less output, has higher costs, and sells its output for a higher price than it would if constrained by competition. These negative outcomes usually generate government regulation.

natural monopoly: An industry in which one large producer can produce output at a lower cost than many small producers. It undersells its rivals and ends up as the only firm surviving in its industry.

nominal interest rates: Interest rates that measure the returns to a loan in terms of money borrowed and money returned (as opposed to *real interest rates*).

nominal prices: Money prices, which can change over time due to inflation. (See also *real prices*.)

nominal wages: Wages measured in money. (See also *real wages*.)

oligopoly: An industry with only a few firms. If they collude, they form a *cartel* to reduce output and drive up profits the way a *monopoly* does.

open-market operations: The buying and selling of government bonds by a central bank; that is, transactions that take place in the public, or open, bond market.

opportunity cost: The value of the next best alternative thing that you could have done. It measures what you gave up in order to do the most preferred thing.

perfect competition: A situation where numerous small firms producing identical products compete against each other in a given industry. Perfect competition leads to firms producing the *socially optimal output level* at the minimum possible cost per unit.

price ceiling: A market intervention in which the government ensures that the price of a good or service stays below the free market price.

price floor: A market intervention in which the government keeps the price of a good or service above its free market price.

Prisoners' Dilemma: A situation in which a pair of prisoners (or firms) has to decide whether or not to co-operate. The dilemma is that although the individual incentives favour not co-operating, if both players figured out a way to co-operate, they'd be better off.

producer surplus: The gain that producers receive when they can sell their output at a price higher than the minimum amount for which they are willing to make it.

Production Possibilities Frontier (PPF): A graph economists use to help them visualise the trade-offs you make when you efficiently reallocate inputs from producing one thing to producing another; sometimes referred to as the *Production Possibilities Curve*.

productive efficiency: A term describing firms that produce goods and services at the lowest possible cost.

public goods: Goods or services that can't be profitably produced by private firms because they are impossible to provide to just one person; if you provide them to one person, you have to provide them to everybody. Because all consumers hope that somebody else is going to pay for public goods so that they can get them for free, nobody ends up paying. Public goods are defined as being non-excludable (you can't prevent anyone from consuming them) and non-rival (it costs no extra to supply one extra person).

quantity demanded: How much of a good or service a consumer demands at a specific price given his or her income and preferences.

quantity theory of money: The theory that the overall level of prices in the economy is proportional to the quantity of money circulating in the economy.

rational expectations: The theory that people optimally change their behaviour in response to policy changes. Depending on the situation, their behavioural changes can greatly limit the effectiveness of policy changes.

real interest rates: Interest rates that compensate for inflation by measuring the returns to a loan in terms of units of stuff lent and units of stuff returned (as opposed to *nominal interest rates*).

real prices: How much of one kind of thing (such as hours worked) you have to give up to get a good or service, no matter what happens to *nominal prices*.

real wages: Wages measured not in terms of money itself (as *nominal wages* are) but rather in terms of how much output that money can buy.

recessions: Periods of time in a business cycle during which an economy's total output falls.

recoveries: Periods of time in a business cycle during which an economy's total output expands.

scarcity: The fact that people don't have enough resources to satisfy all their wants; the phenomenon that creates the need for *economics*.

short-run shutdown condition: A situation in which a firm's total revenues are less than its *variable costs*, and the firm is better off shutting down immediately and losing only its *fixed costs*.

socially optimal output level: The output level that maximises the benefits that society can get from its limited supply of resources.

sticky prices: Prices that are slow to adjust to shocks. Price stickiness can cause recessions to linger.

supply and demand: An economic model of markets that separates buyers from sellers and then summarises each group's behaviour with a single line on a graph. The buyers' behaviour is captured by the *demand curve*, whereas the sellers' behaviour is captured by the *supply curve*. By putting these two curves on the same graph, economists can show how buyers and sellers interact in markets to determine how much of any particular item is going to be sold, as well as the price at which it's going to be sold.

supply curve: A line on a graph that represents how much of a good or service sellers are going to produce at various prices.

total surplus: The sum of *producer surplus* and *consumer surplus*.

traditional economy: An economy in which production and distribution are handled along the lines of long-standing cultural traditions.

Tragedy of the Commons: If a resource is open to public use, it typically becomes rapidly exhausted or ruined because each person's personal incentive is to use it up before anyone else can. This problem is solved by private property rights, which give owners an incentive to conserve the resource and harvest it at sustainable rates.

utility: A measure of happiness that economists suppose people use to compare all possible things that they may experience.

variable costs: Costs that vary with the amount of output produced.

Index

• A •

absolute advantage, 86–87
abstracting, 19
accounting profit, 222–223
actual expenditure, 132–133, 135
actual investment, 133
adverse selection, 337–338
AFC (average fixed cost), 227
aggregate demand
 adjusting to shift in, 119–120
 defined, 369
 fiscal policy, 151
 negative and positive demand shock, 119–120
 results of increasing, 145–146
aggregate demand (AD) curve
 inverse relationship, 117
 shift, adjusting to recession, 144
aggregate shift, 115
aggregate supply, 369
Akerlof, George (economist)
 Nobel Prize in Economics, 331
alcohol and tobacco, 54
allocation
 allocative efficiency, 369
 inflation, 210
 optimal, 44
 resource, 43–44, 46
allocative efficiency, 40
allocatively efficient profit-maximising firm, 17
altruism, 28
American Telephone and Telegraph (AT&T) Corporation, 289–290
ancient living standards, 10
antitrust law, 306–307, 369
Arrow, Kenneth (economist), 350

asset
 flow, 70
 gross domestic product (GDP), 69–70
 international trade, 83–84
asymmetric information
 described, 18, 329
 insurance, 335–339
 trade, limiting, 330–331
 used car market, 331–335
ATC (average total cost), 228–231
AT&T (American Telephone and Telegraph) Corporation, 289–290
autonomous expenditure, 134
AVC (average variable cost), 226, 229
average cost
 confusing, 37
 marginal versus, 37
 pricing, 288
average fixed cost (AFC), 227
average total cost (ATC), 228–231
average variable cost (AVG), 226, 229

• B •

balanced budget, 79
balanced technological change, 48
bankruptcy, 333
barter system, 91
Becker, Gary (economist), 352–353
behaviour, consumer
 diminishing marginal utility, 202–206
 final choice, making, 32–33
 happiness, maximising, 27–29
 human behaviour modeling, 26
 irrationality, 35–37
 limitations, 29–32
 marginal utility, 32–33
 scarcity, 25
 uninformed decision-making, 34–35
Bentham, Jeremy (philosopher), 199

biased technology innovation, 48–49
 bi-metallic (gold and silver) standard, 97
 bond
 coupon payment, 160
 monetary policy, 160
 open-market operation, 163
 price link to interest rate, 161
 zero-coupon, 160
 borrower, 96
 borrowing, 153
 breaking even, 222
 British Telecom, 289
 Bryant, William Jennings (Nebraska senator), 97
 budget deficit, 79, 153
 budget surplus, 79
 Burger King restaurant, 313
 business cycle, 112–113

• C •

c (consumption expenditure) equation, 75–77
 cable television monopoly, 284
 CAP (Common Agricultural Policy), 56
 capital
 defined, 369
 as factor of production, 41
 human, 41–42
Capital (Marx), 348
 capital stock, 77
 carbon emission, 327
 cardinal utility, 199
 Carlyle, Thomas (English historian), 171
 cars, market for used
 asymmetric information, 331–335
 credit crunch, 331–335
 expert opinion, getting, 335
 poor quality, reasons for, 331–333
 reputation building, 334–335
 warranty, offering, 334
 cartel. *See also* Prisoner's Dilemma
 collusion difficulty, 295–296
 coordinating, 294
 defined, 369
 OPEC oil, 295–296
 Prisoner's Dilemma, applying, 302–306
 trusts, 306–307

causation, 357
 centralisation, 53
 child labour, 54
 collusion
 competition outcome comparison, 293–294
 difficulty of, 295–296
 outcome, 292–293
 command economy, 53, 57, 369
 command land, 324–326
 commodity, 220
 Common Agricultural Policy (CAP), 56
 common good, self-interest promoting, 28–29
 Commons, Tragedy of the
 defined, 375
 environmental problems and poor property rights, 326–327
 overgrazing commonly owned field, 324–326
 Communism, 53, 58, 313
 comparative advantage, 85–87, 369
 competition. *See also* monopolistic competition
 collusion outcome comparison, 293–294
 competitive firm and monopoly comparison, 280–283
 imperfect, 219
 invisible hand theory, 17
 lack of, problems with, 17–18
 market price, 241
 microeconomics, 17
 monopolistic, 219
 monopoly, breaking into several competing firms, 289–290
 perfect, 219–220, 260
 profit maximisation firm, 219–223
 redundant, 284
 competitive free market. *See also* free market
 benefits, 260
 causes and consequences, 260–261
 process, 261–262
 profits guiding firm entry and exit, 263–267
 socially optimal output, 244
 competitive market, 51–52, 369
 complementary good, 215

- composition fallacy, 358–359
 confession, in Prisoner's Dilemma, 297–298
 consequence, unintended, 362
 constrained maximisation, 351
 constrained optimisation, 198
 constraint
 resource, 29–30
 technology, 30
 time, 30–31
 consumer choice
 behaviour and scarcity, 25
 diminishing marginal utility, 202–206
 final choice, making, 32–33
 happiness, maximising, 27–29
 human behaviour modeling, 26
 irrationality, 35–37
 limitations, 29–32
 marginal utility, 32–33
 uninformed decision-making, 34–35
 Consumer Price Index (CPI), 101, 105, 370
 consumer surplus
 of continuous good, 251–252
 defined, 370
 of discrete good, 249–250
 consumption expenditure (C) equation, 75–77
 consumption function, 133
 continuous good, 251–252
 copyright, 60
 Corn Laws, 85
 cost
 average fixed cost (AFC), 227
 average total cost (ATC), 228–231
 average variable cost (AVG), 226, 229
 fixed, 224, 371
 imposed on others, failing to take account of, 318–320
 per unit of output, 224–226
 sunk, 35–36
 variable, 224
 cost-benefit analysis
 marginal utility, 32–33
 total utility, 33
 cost change, supply curve, 182–183
 cost structure analysis, 223–225
 coupon payment, 160
 CPI (Consumer Price Index), 101, 105, 370
 credit crunch, 331–335
 Croesus, King of Lydia, 93
 cross-price effect, 214–215
- D •
- deadweight loss
 defined, 370
 monopoly, 282
 price ceiling, dissecting from, 255–256
 of tax, 256–259
 Debreu, Gerard (economist), 350
 debt
 national, 153
 paying, by printing money, 154–155
 rolling over the, 154
 decision-making, uninformed, 34–35
 deferred payment standard, 96, 99
 deficit
 budget, 79, 153
 fiscal policy, 153–155
 trade, 81–82
 deflation, 370
 demand. *See also* supply and demand
 defined, 370
 described, 172
 downward-sloping, 309–310
 elasticity, 178–180
 facing down, 271–272
 increase, reacting to, 189–190
 inferior good, 173
 inverse relationship, 173
 Law of Demand, 20
 normal good, 173
 quantity demanded, 173
 shift in, 175–176
 stimulating to end recession, 142–144
 demand curve
 cross-price effect, 214–215
 data, plotting, 20–21
 defined, 370
 deriving marginal revenue from, 272–274
 described, 19–20
 diminishing marginal utility, 211–215
 drawing, 22–23
 graphing, 174–176
 higher output and lower prices, 275–276
 inverse relationship, 20
 marker, 220

- demand curve (*continued*)
- opportunity cost, 176–178
 - own-price effect, 214
 - price change, 174–175
 - shift, 175–176, 213
 - slope, determining, 176–178
- democracy, 11, 364
- depreciation, 77, 370
- depression, 128
- diminishing marginal utility. *See also* marginal utility; utility-maximising
- consumer
 - allowing for, 34
 - consumer choice, 202–206
 - defined, 32, 370
 - demand curve, 211–215
 - getting less from more, 199–202
 - income effect, 213
 - two goods, allocating money between, 205–208
- diminishing return
- cause, 226
 - defined, 370
 - determining what is possible to produce, 40–43
 - input/output combinations, 48
 - low-hanging fruit principle, 42
 - output, increasing cost for producing, 43
 - scarcity, 14
 - technology shift, 48
 - unavoidable nature of, 14
- discrete good, 249–250
- discrimination, 339
- domestically made goods and services, 80
- dominant firm, 306
- dominant strategy, in Prisoner’s Dilemma, 299–300
- downward price stickiness, 149
- downward-sloping demand, 309–310
- downward wage stickiness, 149
- duopoly, 296
- **E** ●
- economic cost, 370
- economic model, 19, 370
- economic profit, 222–223, 370
- economics
- defined, 9, 370
 - welfare, 349
- economist disagreement, 2
- economy
- command, 53, 57, 369
 - laissez-faire, 57–58
 - market, 57, 374
 - measuring, 14–15
 - traditional, 57
- economy, stimulating
- to end recession, 142–144
 - failure to stimulate, 149–151
 - with inflation, 96–98
 - interest rate lowering, 163–164
 - risk to too much, 144–148
- education
- government encourage, 61
 - living standards, 12, 364
- efficiency
- allocative, 369
 - free market, 246, 249
 - monopoly, 283
 - productive, 40
- elasticity, demand, 178–180
- electromagnetic spectrum, 184
- environmental problems, 326–327, 364–365
- equation
- Fisher, 107
 - inventory adjustment, 133–135
- equation, GDP
- C (consumption expenditures), 75–77
 - described, 74
 - G (government purchases of goods and services), 75, 78–80
 - I (investment expenditures), 75, 77–78
 - NX (net exports), 75, 80–81
- equilibrium
- adjusting to new, 189–192
 - constructing impediments to, 192–195
 - finding, 184–186
 - firm entry and exit, 311–312
 - graphing, 136–138
 - interacting supply and demand to find, 185–188
 - level of price, 117
 - monetary policy, 159

price, 186
 price ceiling, 192–193
 price floor, 193–194
 price reduction until reached, 187–188
 quantity, 186
 raising prices until reached, 188
 stability, 187–188
 stable, 136, 159
 EX (exports), 80
 expenditure
 actual, 132–133, 135
 autonomous, 134
 induced, 134
 planned, 132–135
 exports (EX), 80
 externality
 cost and benefit others feel from our
 actions, 317–321
 defined, 370
 negative, 318–322
 positive, 318, 322–324

● F ●

face value payment, 160
 factors of production, 41, 371
 fallacy
 composition, 358–359
 foreign wages, low, 360–361
 free market, unstable, 360
 labour, lump of, 355–356
 overpopulation, 356–357
 protectionism, 357–358
 sequence, confusing with causation, 357
 tax rate and work effort, 361
 unintended consequence, 362
 farming, 97
 Federal Reserve Bank, 105
 fiat money
 adoption of, 157
 benefit, 156–157
 described, 95
 metallic standard, 157
 money supply, expanding the, 156
 fiat system, 95
 financial markets
 defined, 371
 gross domestic product (GDP), 71

firm
 allocatively efficient, 17
 attracting new, 262
 cost structure, 223–226
 dominant, 306
 entry and exit, 263–267
 limited liability corporation, 11
 losing existing, 262
 loss, visualisation, 237–238
 maximising profit for, 281–282
 monopoly, 219
 oligopoly, 292–293
 ownership, 67–68
 productively efficient, 17
 profit, visualisation, 234–236
 firm, profit maximisation
 average fixed cost (AFC), 227
 average total cost (ATC), 228–231
 average variable cost (AVC), 226–227
 competition, 219–223
 cost per unit of output, 224–226
 cost structure analysis, 223–225
 described, 217
 long-run shutdown condition, 240
 loss, visualising, 237–238
 marginal cost, 229
 marginal revenue with marginal cost
 comparison, 231–234
 market price, 241–242
 monopoly, 219, 270
 output level to, 276–280
 profit, visualising, 234–236
 reasons to study, 217–218
 short-run shutdown condition, 238–240
 fiscal policy
 aggregate demand, 151
 deficit, 153–155
 defined, 371
 described, 16
 future tax revenue reliance, 154
 government intervention, 141
 government spending, increasing,
 152–153
 macroeconomics, 16
 recessions, fighting, 16
 Fisher, Irving (Fisher equation), 107
 fishing industry, 326–327
 fixed cost, 224, 371

fixed price, 120–123
 flow, asset, 70
 foreign wage, 360–361
 free enterprise, 366–367
 free market. *See also* competitive free market
 benefits, 244–246
 consumer surplus, 249–252
 cost and benefit comparison, 246–248
 deadweight loss, 255–260
 described, 243
 efficiency of, 246–249
 gain, total surplus measuring, 249–253
 market failure, 243
 prerequisites, 244–246
 price-taking assumption, 245
 producer surplus, 252–253
 regulation, 364
 socially optimal quantity of output, 244, 248–249
 total surplus computing, 253–254
 unstable fallacy, 360
 free trade, 84, 366
 freedom and democracy, 364
 frictional unemployment, 114
 Friedman, Milton (economist), 351
 full-employment output
 aggregate demand, results of increasing, 145–146
 defined, 371
 demand, stimulating, 142–144
 frictional unemployment, 114
 versus maximum output, 114
 recession, 113–114, 142–144
 striving for, 113–114
 future
 inflation expectation/prediction, 106–109
 resources, uses of, 12

● G ●

G (government purchases of goods and services) equation, 75, 78–79, 90
 gain, 249–253
 game theory, 296

GATT (General Agreement of Tariffs and Trade), 84
 GDP (gross domestic product)
 accounting convention, 72
 assets, 69–70
 circular flow diagram, 68, 71
 defined, 14, 371
 equation, 74–81
 exclusions from, 67
 financial markets, 71
 higher, 73
 importance of calculating, 15, 67
 income, 68–69
 macroeconomic measurement, 81–87
 markets for factors of production, 71
 newly produced output counted as, 72
 sticky price equilibrium, 138–139
 General Agreement of Tariffs and Trade (GATT), 84
 general price, 100
The General Theory of Employment, Interest and Money (Keynes), 74
 generosity, 28
 gold and silver (bi-metallic) standard, 97
 gold standard, 94–95
 good
 complementary, 214
 continuous, 251–252
 discrete, 249–250
 inferior, 173
 normal, 173
 substitute, 214
 goods and services
 domestically made, 80
 government intervention, case for, 55
 marginal utility, 208–211
 markets for, 71
 negative consequence, 51
 subsidies to encourage production of, 55
 unequal distribution, 51
 government bond
 coupon payment, 160
 face value payment, 160
 monetary policy, 160
 open-market operation, 163

price link to interest rate, 161
 zero-coupon, 160

government intervention
 alcohol and tobacco, 54
 case against, 55
 case for, 54–55
 education, encouraging, 61
 fiscal policy, 141
 market, 51–56
 monetary policy, 141
 monopoly, 269
 patents, 60
 price system, 55
 production process, 49–56
 pros and cons, 50–51
 sticky prices, 126–127
 technology and innovation, encouraging,
 59–61

government purchases of goods and
 services (G) equation, 75, 78–80

government spending
 borrowing, 153
 increasing, 152–153

graphing. *See also* model
 abstracts, 19
 demand, 19–22
 demand curve, 20–23, 174–176
 equilibrium, 136–138
 inventory adjustment, 136–138
 money supply increase results, 164–166
 Production Possibilities Frontier (PPF),
 44–47
 supply curve, 180–183

Great Depression, 74, 127–128, 157

Greece, innovation in, 11

gross domestic product. *See* GDP

group discrimination, 339

• H •

happiness, maximising
 altruism and generosity, 28
 desire to be happy, 27
 self-interest promoting common good,
 28–29
 utility measuring, 27

history
 ancient living standards, 10
 hyperinflation, 94
 inflation, 91
 institutions that improved living
 standards, 11–12
 Standard Oil Company, 306

Hitler, Adolf (German government), 94

Hope Diamond, 184

household
 consumption, 75–76
 firms, ownership of, 67–68

human behaviour modeling, 26

human capital, 41–42, 371

hyperinflation. *See also* inflation
 defined, 371
 gold standard, 94–95
 in history, 94
 money, loss of effectiveness, 95
 occurrence of, 89
 printing money, 94–96

• I •

I (investment expenditure) equation,
 75, 77–78

illegal drug, 54

IM (imports), 80

imperfect competition, 219

income
 flow of, 69
 gross domestic product (GDP), 68–69
 income and wealth inequality, 54
 income effect, 213
 increasing return, 225, 371
 induced expenditure, 134
 inefficient production, 312
 inferior good, 173

inflation. *See also* hyperinflation
 allocation, 210
 barter system, 91
 causes, 89
 defined, 371
 effects, tallying, 98–100
 failure to stimulate, 149–151
 farming, 97

- inflation (*continued*)
 - fiat system, 95
 - future expectation/prediction, 106–109
 - history, 91
 - importance of calculating, 15
 - money supply and demand, balancing, 90–92
 - nominal wage, raising while prices are stuck, 147–148
 - prevention, 367
 - price index measurement, 100–104
 - rate of, calculating, 102
 - rational expectation, 166
 - real wages, tracing movement of, 146–148
 - risk of too much stimulation, 144–148
 - stagflation, 167
 - stimulating economy with, 96–98
 - targets, 167–168
 - tax, 99–100
 - temptation, giving into, 92–98
- inflation rate, 371
- innovation
 - biased-technology, 48–49
 - economic growth dependent on, 364
 - patent, 284
 - production, encouraging, 59–61
- An inquiry into the Nature and Causes of the Wealth of Nation* (Smith), 28
- institution, living standards, 11–12
- insurance
 - adverse selection, avoiding, 337–338
 - asymmetric information, 335–339
 - grouping individuals, 336–337
 - moral hazard, migrating, 338–339
- intellectual property rights, 364
- interest rate
 - capital stock, 77
 - changing money supply to change, 162–163
 - defined, 371
 - investment relationship, 78
 - lowering, 163–164
 - monetary policy, 162–163
 - nominal, 106–107, 373
 - price link to, 161
 - real, 106–107
 - supply and demand for loans
 - determining, 71
- International Organisation for Economic Co-operation and Development, 78
- international trade
 - asset, 83–84
 - comparative advantage, 85–87
 - free trade, 366
 - impacts on economy, 81–87
 - macroeconomic measurement, 81–87
- inventory
 - graphing, 136–138
 - model, 129–130
- inventory adjustment
 - actual expenditure, 132–133
 - autonomous expenditure, 134
 - consumption function, 133
 - equation, 133–135
 - induced expenditure, 134
 - novelty of, 129–130
 - output, increasing or decreasing, 131
 - planned expenditure, 132–133
 - target level, 130–131
- inverse relationship
 - aggregate demand (AD) curve, 117
 - demand, 173
 - demand curve, 20
 - money value and prices, 91
 - quantity demanded and price, 20–21
- investment
 - actual, 133
 - in capital stock, 77–78
 - defined, 371
 - planned, 133
- investment expenditure (I) equation, 75, 77–78
- invisible hand theory
 - competitive markets, 17
 - defined, 371
 - production output, 265
 - property rights, 315
 - self-interest, promoting the common good, 29, 347

irrationality
 consumer choice, 35–37
 marginal and average cost, confusing, 36–37
 percentages and dollar amounts, confusing, 36
 sunk cost, 35–36

• K •

Keynes, John Maynard (economist)
 anti-recession policies, 15
 equilibrium with sticky prices, 127–138
The General Theory of Employment, Interest and Money, 74
 Keynesianism, 129–130
 modern macroeconomics, 350
 Khan, Kublai, 93

• L •

labour
 child, 54
 cutting wages versus cutting workers, 125
 as factor of production, 41
 flow of services needed in production, 69
 lump of labour fallacy, 355–356
 misdeeds of market, 54
 laissez-faire economy, 57–58, 130
 land
 common, 324–326
 as factor of production, 41
 as perfectly inelastic supply, 184
 law
 antitrust, 306–307, 369
 negative externality, 321
 Law of Demand, 20, 371
The Law of Unintended Consequences, 362
 legal monopoly, 162
 Lehman Brothers' bankruptcy filing, 333
 life expectancy, 10
 limitation
 opportunity cost, 29, 31–32
 resource constraint, 29–30

technology constraint, 30
 time constraint, 30–31
 limited liability corporation, 11
 limited resource, 40. *See also* scarcity
 Littlefield, Henry (professor), 97
 living standards
 ancient, 10
 education, 12, 364
 institutions that improved, 11–12
 price index, determining real, 104–105
 long-run aggregate supply curve (LRAS), 117–119
 long run recession effect
 described, 115
 negative demand shock, response to, 123–124
 price adjustment, 118–119
 versus short run, 116
 long-run shutdown condition, 240–241, 371
 loss
 deadweight, 255–260
 visualization, 237–238
 low-hanging fruit principle, 42
 LRAS (long-run aggregate supply curve), 117–119
 Lucas, Robert (economist), 353

• M •

macroeconomic measurement
 gross domestic product (GDP), 66–71, 73
 international trade, 81–87
 National Accounts accounting apparatus, 65
 reasons for, 65
 macroeconomics
 conventions used in the book, 5
 defined, 13, 372
 household consumption, 75–76
 measuring economy, 14–15
 model, 116
 monetary and fiscal policies, 15–16
 recession, causes of, 15

- marginal cost
 - average, confusing, 36–37
 - defined, 372
 - focusing on, 229
 - marginal revenue comparison, 231–234, 236–237
 - private, 319–320, 322
 - social, 319–320, 322
- marginal cost pricing, 288
- marginal output, 225
- marginal propensity to consume (MPC), 76
- marginal revenue
 - decreasing while increasing production, 274–275
 - deriving from demand curve, 272–274
 - higher output and lower prices, 275–276
 - marginal cost comparison, 231–234, 236–237
 - relating to total revenue, 274
- marginal utility. *See also* diminishing marginal utility
 - buying as much as you can, 203–206
 - cost-benefit analysis, 32–33
 - defined, 372
 - diminishing, 32, 34
 - goods and services, 208–211
- marginalism, 359
- market
 - commodity, 220
 - competitive, 51–52
 - computerised, 172
 - defined, 372
 - demand curve, 220
 - for factors of production, 71
 - for goods and services, 71
 - government intervention, 51–56
 - institutional arrangement governing, 172
 - misdeeds of, 53–54
 - open-market operation, 163, 373
 - people wanting to participate in, 245
 - price system, 52
 - supply and demand, 172
 - supply curve, 220
- market, free
 - benefits, 244–246
 - consumer surplus, 249–252
 - cost and benefit comparison, 246–248
 - deadweight loss, 255–260
 - described, 243
 - efficiency of, 246–249
 - gain, total surplus measuring, 249–253
 - market failure, 243
 - prerequisites, 244–246
 - price-taking assumption, 245
 - producer surplus, 252–253
 - regulation, 364
 - socially optimal quantity of output, 244, 248–249
 - total surplus computing, 253–254
 - unstable fallacy, 360
- market basket, 372
- market basket price index, 100–102, 105
- market economy, 57, 372
- market equilibrium
 - adjusting to new, 189–192
 - constructing impediments to, 192–195
 - equilibrium price, 186
 - finding, 184–186
 - interacting supply and demand to find, 185–188
 - price ceiling, 192–193
 - price floor, 193–194
 - price reduction until reached, 187–188
 - quantity, 186
 - raising prices until reached, 188
 - stability, 187–188
- market failure
 - defined, 372
 - free market, 243
 - microeconomics, 18
 - public goods, 18, 339–344
- market failure, asymmetric information
 - described, 329
 - insurance, 335–339
 - trade, limiting, 330–331
 - used car market, 331–335
- market price, 186, 241
- market production, 51, 372
- market quantity, 186
- Marshall, Alfred (economist), 349–350
- Marx, Karl (economist), 348–349
- maximising happiness
 - altruism and generosity, 28
 - desire to be happy, 27

- self-interest promoting common good, 28–29
- utility measuring, 27
- maximum output, 114
- McDonald's restaurant, 313
- measuring economy, 14–15
- medium of exchange, 99
- mercantilism, 348
- metallic standard, 157
- microeconomics
 - competition, 17
 - conventions used in this book, 5
 - defined, 13, 372
 - failures, common market, 18
 - household consumption, 75
 - problems without competition, 17–18
 - property rights, 18
 - scope of, 16
 - supply and demand, balancing, 16–17
- mixed economy production, 56–59
- model. *See also* graphing
 - abstracting, 19
 - consumer choice, 26
 - demand curve, 19–22
 - inventory, 129–130
 - macroeconomic, 116
- monetary policy
 - defined, 156, 372
 - described, 15
 - fiat money benefits, 156–157
 - government bond, 160
 - government intervention, 141
 - inflation, stimulating economy with, 96
 - interest rate change, 162–163
 - macroeconomics, 15
 - money, realisation of too much, 158–159
 - open-market operation, 163
 - printing money, 96
 - rational expectation, 158–159, 164–168, 353
 - recessions, fighting, 15
 - stable equilibrium, 159
- money
 - fiat, 95, 156–157
 - inflation, 90–92
 - loss of effectiveness, 95
 - medium of exchange, 99
 - printing, 94–96
 - quantity theory of, 92
 - realisation of too much, 158–159
 - standard of deferred payment, 99
 - store of value, 98
 - supply and demand, 90–92
 - supply, expanding the, 156
 - supply increase, graphing results of, 164–166
 - unit of account, 98
- monopolistic competition
 - basic description of, 292
 - defined, 372
 - facing competition, 219
 - product differentiation, 307–308
 - profit limit, facing, 308–313
- monopoly. *See also* oligopoly
 - breaking into several competing firms, 289–290
 - cable television, 284
 - competitive firm comparison, 280–283
 - deadweight loss, 282
 - decreasing marginal revenue, 271–276
 - defined, 17, 373
 - demand curve, deriving marginal revenue from, 272–274
 - efficiency, 283
 - facing down demand, 271–272
 - good examples of, 283–285
 - government intervention, 269
 - harm caused by, quantifying, 282
 - legal, 162
 - less production of, 281–282
 - minimum output requirement, 286–287
 - natural, 285
 - natural gas, 284
 - non guarantee of profit, 278–279
 - oligopoly, 219
 - output and price level, 280
 - output level maximising profit, 276–280
 - patent, encouraging innovation and investment with, 284
 - pricing regulation, 287–289
 - problems caused by, 270
 - profit, 278–279
 - profit maximisation firm, 219, 270
 - redundant competition, 284
 - regulation, 285–289

monopoly (*continued*)
 rubbish collection, 284
 subsidizing to increase output, 286
 moral hazard, 338–339
 MPC (marginal propensity to consume), 76

• N •

National Accounts accounting apparatus, 65
 National Bureau of Economic Research (NBER), 129
 national debt, 153
 National Income and Product Accounts (NIPA), 65
 natural gas monopoly, 284
 natural monopoly, 285, 373
 NBER (National Bureau of Economic Research), 129
 negative demand shock
 aggregate demand, adjusting to shift in, 119–120
 demand, full-employment output, 142
 long-run and short-run response, 123–124
 negative externality
 cost imposed on others, failing to take account of, 318–320
 dealing with, 321–322
 defined, 318
 effects of, 318–320
 positive amount of, 320–321
 things that impose, overproducing, 319–320
 neoclassical synthesis, 351
 net export (NX) equation, 75, 132
 NIPA (National Income and Product Accounts), 65
 Nixon, Richard (President), 95
 Nobel Prize in Economics
 Akerlof, 331
 Ostrom, 326
 nominal interest rate, 106–107, 373
 nominal price, 104, 373
 nominal wage, 147–148
 non-excludable public good, 340
 non-rival public good, 339

normal good, 173
 NX (net export) equation, 75, 80–81, 132

• O •

OECD (Organisation for Economic Cooperation and Development), 101, 129
 oil production, 305
 oligopoly, 219. *See also* monopoly
 antitrust law, 306–307
 basic description of, 291–292
 collusion and competition outcome, 292–293
 defined, 373
 dominant firm, 306
 illegal price fixing, 18
 regulation, 306–307
 strategic interaction, 292–293
 omerta system, in Prisoner's Dilemma, 300–302
 OPEC (Organisation of the Petroleum Exporting Countries) cartel
 collusion, difficulty of, 295–296
 trapped in Prisoner's Dilemma, 304–305
 open-market operation, 163, 373
 opportunity cost
 accounting and economic profit, 222–223
 defined, 373
 demand curve, 176–178
 as limitation, 29, 31–32
 PPF graph, 46
 resource allocation, 46
 optimal allocation, 44
 ordinal utility, 199
 Organisation for Economic Cooperation and Development (OECD), 101, 129
 Organisation of the Petroleum Exporting Countries (OPEC) cartel
 collusion, difficulty of, 295–296
 trapped in Prisoner's Dilemma, 304–305
 Ostrom, Elinor (Nobel Prize in Economics), 326
 overpopulation, 356–357
 own-price effect, 214
 ownership, firm, 67–68

● p ●

- paper currency, 94–96
- patent
 - encouraging innovation and investment with, 284
 - full-time inventors, 12
 - government intervention, 60
 - technology that can't be, 343
 - to turn public good into private good, 343
- pay-off matrix, in Prisoner's Dilemma, 298–299
- payment
 - coupon, 160
 - deferred payment standard, 96, 99
 - face value, 160
 - transfer, 78–79
- perfect competition
 - cause and consequence, 260–261
 - commodity, 220
 - defined, 373
 - market demand curve, 220
 - market supply curve, 220
 - product differentiation, 292
 - requirement, 219–220
- perfectly inelastic supply, 184–185
- philanthropy, 341
- planned expenditure, 132–134
- planned investment, 133
- positive demand shock, 119–120
- positive externality
 - consequence, 322–324
 - defined, 318
 - subsiding things that provide, 324
 - unproducing things that provide, 323–324
- Post hoc ergo propter hoc* phrase, 357
- poverty, 11–12
- PPF (Production Possibilities Frontier)
 - defined, 374
 - inefficient production technology/poor management, 47
 - limited resource, 44
 - opportunity cost or production, 46
 - output combinations, 45
 - scarce resource of production, 46
- prediction, inflation, 106–109
- price
 - cross-price effect, 214–215
 - downward price stickiness, 149
 - elasticity of demand, 178–180
 - equilibrium, 186
 - equilibrium level of, 117
 - fixed, 120–123
 - general, 100
 - holding constant, 175
 - marginal cost, 288
 - market, 186
 - money, value of, 91–92
 - monopoly regulation, 287–289
 - nominal, 104, 373
 - own-price effect, 214
 - real, 104
 - sales and production cost, separating, 181
 - upward price stickiness, 149
- price adjustment
 - inventory adjustment instead of, 129–130
 - long run, 118–119
 - natural result of, 114–115
- price ceiling
 - defined, 373
 - dissecting deadweight loss from, 255–256
 - market equilibrium impediment, 192–193
- price change
 - demand curve, 174–175
 - quantity demanded affect, 211–213
 - supply curve, 181–182
 - what price to change, 278
- price floor, 193–194, 373
- price index
 - Consumer Price Index (CPI), 101, 105, 370
 - general price, 100
 - market basket, 100–102, 105
 - price level index, 103
 - problems, 105–106
 - rate of inflation calculation, 102
 - relative price, 100
 - Retail Price Index (RPI-X), 101
 - setting up, 103
 - standard of living, determining real, 104–105
 - Undergraduate Price Index, 101–102

- price system
 - government intervention lacking
 - flexibility of, 55
 - resource allocation, 52
- price taker, 219–223, 260
- price-taking assumption, 245
- printing money
 - hyperinflation, heading for, 94–95
 - monetary policy, 96
 - paying debt by, 154–155
- printing press, 96, 342
- Prisoner's Dilemma
 - cartels, 302–306
 - confessing or remaining silent, 297–298
 - defined, 373
 - dominant strategy, determining, 299–300
 - duopoly, 296
 - game theory, 296
 - lousy outcome, realising, 300
 - omerta system to resolve, 300–302
 - pay-off matrix, 298–299
- private marginal cost, 319–320, 322
- producer surplus
 - defined, 374
 - measurement, 252–253
- product differentiation
 - benefit, 307–308
 - Communism, 313
 - perfect competition, 292
- production
 - capital as factor of, 41
 - factors of, 41
 - flow of services needed in, 69
 - government intervention, 50–56
 - inefficient, 312
 - labour as factor of, 41
 - land as factor of, 41
 - market, 51
 - markets for factors of, 71
 - mixed economy, 56–59
 - oil, 305
 - possibilities, determining, 39–49
 - productive versus allocative efficiency, 49–50
 - resource allocation, 43–44
 - resource used in, classifying, 41–42
 - technology and intervention,
 - encouraging, 59–61
 - traditional, 56
- Production Possibilities Frontier (PPF)
 - defined, 374
 - inefficient production technology/poor management, 47
 - limited resource, 44
 - opportunity cost or production, 46
 - output combinations, 45
 - scarce resource of production, 46
- productive efficiency, 40, 374
- productively efficient profit-maximising firm, 17
- profit
 - accounting, 222–223
 - economic, 222–223
 - monopoly, 278–279
 - running a, 222
 - as self-correcting feedback mechanism, 265
 - sticky prices, 126
 - visualisation, 234–236
- profit limit
 - downward-sloping demand, 309–310
 - facing, 308–313
 - firm entry and exit, finding equilibrium, 311–312
 - inefficient production, 312
- profit maximisation
 - average fixed cost (AFC), 227
 - average total cost (ATC), 228–231
 - average variable cost (AVC), 226–227
 - competition, 219–223
 - cost per unit of output, 224–226
 - cost structure analysis, 223–225
 - described, 217
 - long-run shutdown condition, 240
 - loss, visualising, 237–238
 - marginal cost, 229
 - marginal revenue with marginal cost comparison, 231–234
 - market price, 241–242
 - monopoly, 219, 270
 - output level to, 276–280
 - profit, visualising, 234–236
 - reasons to study, 217–218
 - short-run shutdown condition, 238–240
- property rights
 - bad outcome, 315
 - common land, 324–326
 - cost and benefit others feel from our actions, 317–321

environmental problem, 364–365
 extinction caused by, 326–327
 full and complete, 317
 intellectual, 364
 invisible hand theory, 315
 microeconomics, 18
 negative externality, 318–322
 poor, 326–327
 positive externality, 318, 322–324
 protection, 364
 socially optimal outcome, 316–317
 Tragedy of the Commons, 324–327
 protectionism, 357–358
 public good
 defined, 374
 described, 18, 330
 free enterprise, 366–367
 non-excludable, 340
 non-rival, 339
 philanthropy to provide, 341
 private good, selling related, 341–342
 tax, 340–341
 technology as, 342–344

• **Q** •

quality
 income and wealth inequality, 54
 used car market, 331–333
 quantity demanded
 defined, 374
 inverse relationship to price, 20–21
 quantity theory of money, 92, 374

• **R** •

rational expectation
 defined, 374
 inflation, 166
 monetary policy, 158–159, 164–168, 353
 price stickiness, 151
 real interest rate
 defined, 374
 prediction, 106–109
 real price, 104, 374
 real wage
 decrease in, 148
 defined, 374
 tracing movement of, 146–148

recession. *See also* fiscal policy; monetary policy
 business cycle, 112–113
 causes of, 15
 defined, 129, 374
 described, 113
 failure to stimulate, 149–151
 fighting with monetary and fiscal policies, 15–16
 full-employment output, 113–114, 142–144
 government spending to help end, 153–154
 inflation, risk of too much stimulation, 144–148
 long-lasting consequences, 111
 long run effects, 115–116, 118–119, 123–124
 price adjustment, 114–115
 recovery, 113
 shock, responding to, 115–117
 short run effects, 115–116, 120–124
 stick price, 124–127
 stimulating demand to end, 142–144
 recovery
 business cycle, 113
 defined, 374
 redundant competition, 284
 regulation
 free market, 364
 monopoly, 285–289
 oligopoly, 306–307
 relative price, 100
 resource
 constraint, 29–30
 future, use of, 12
 production, 41–42
 resource allocation
 opportunity cost, 46
 price system, 52
 production process, 43–44
 Retail Price Index (RPI-X), 101
 Ricardo, David (economist)
 comparative advantage, 85, 348
 mercantilism, 348
 Robinson, Joan (economist), 141, 308
 Rockefeller, John D. (Standard Oil Company), 306
 rolling over the debt, 154
 RPI-X (Retail Price Index), 101

rubbish collection monopoly, 284
 running a profit, 222

• S •

Samuelson, Paul (economist), 351
 Saudi Arabia, 305–306

scarcity

conventions used in this book, 4–5
 defined, 13, 375
 diminishing returns, 14
 as heart of consumer choice, 25
 time and/or physical resources, 14

self-interest

behaviour, analysing, 26
 common good, promoting, 28–29
 trading, 363

shift

aggregate, 115
 in demand, 175–176
 supply curve, 182–183
 technology, 48

shock, negative demand

aggregate demand, adjusting to shift in,
 119–120
 demand, full-employment output, 142
 long-run and short-run response, 123–124

shock, positive demand, 119–120

short-run aggregate supply (SRAS),
 120–121

short run recession effect

described, 115
 fixed prices, dealing with, 120–123
 versus long run, 116
 negative demand shock, response to,
 123–124

short-run shutdown condition, 238–241, 375

silence, in Prisoner's Dilemma, 297–298

Smith, Adam (economist)

importance of, 347–348
 individual happiness, 28–29
*An inquiry into the Nature and Causes of
 the Wealth of Nations*, 28
 invisible hand theory, 17, 29, 265, 315, 347
The Wealth of Nations, 315
 social marginal cost, 319–320, 322

socially optimal output

competitive free market, 244
 defined, 375
 determining the, 248
 free market production, 248–249
 property rights, 316–317

Solow, Robert (economist), 352

special-interest lobbying, 55

spending, government

borrowing, 153
 increasing, 152–153

SRAS (short-run aggregate supply),
 120–121

stable equilibrium, 136, 159

stagflation, 167

standard of deferred payment, 96, 99

Standard Oil Company, 306

statistical discrimination, 339

step function, 249

sticky price

cost of profits and wages, 126
 cutting wages versus cutting workers, 125
 defined, 375

downward price stickiness, 149

downward wage stickiness, 149

government intervention, 126–127

gross domestic product (GDP), 138–139

importance of, 149

rational expectation, 151

recessions, 124–127

upward price stickiness, 149

upward wage stickiness, 149

stimulating economy

to end recession, 142–144
 failure to stimulate, 149–151
 with inflation, 96–98

interest rate lowering, 163–164

risk to too much, 144–148

stimulus

anticipation, 149–151
 rational expectation, 151

stock, 70

strategic situation, 293

substitute good, 214

sunk cost, 35–36

supermarket, 122

- supply. *See also* supply and demand
 aggregate, 369
 decrease, reacting to, 191–192
 perfectly inelastic, 184–185
 supply and demand. *See also* demand;
 supply
 cost and benefit comparison, 246–248
 defined, 375
 described, 171
 inflation, 90–92
 market, 172
 market equilibrium, adjusting to new,
 189–192
 market equilibrium, constructing
 impediments to, 192–195
 market equilibrium, finding, 185–188
 microeconomic balancing, 16–17
 money, 90–92
 supply curve
 cost change, 182–183
 defined, 375
 graphing, 180–183
 market, 220
 price change, 181–182
 sales price and production cost,
 separating, 181
 shift, 182–183
 tax shifting the, 257–258
 surplus
 budget, 79
 consumer, 249–252
 producer, 252–253
 total, 253–254
 trade, 81–82
 sweatshop labour, 54
- **T** ●
- target inventory level, 130–131
 tax
 deadweight loss of, 256–259
 future revenue, relying on security of, 154
 inflation, 99–100
 public good, 340–341
 supply curve shift, 257–258
 tax rate fallacy, 361
 value added tax (VAT), 56
 technology
 balanced technological change, 48
 biased innovation, 48–49
 constraint, 30
 copyright, 60
 diminishing returns, 48
 production, encouraging, 59–61
 as public good, 342–344
 temptation, inflation, 92–98
 Third World suffering, 194
 time constraint, 30–31
 tobacco and alcohol, 54
 total revenue, 274
 total surplus
 computing, 253–254
 contemplation, 254
 defined, 375
 gain measurement, 249–253
 total utility, 33
 trade
 free, 84, 366
 limiting, 330–331
 self interest, 363
 trade, international
 affects on economy, 81–87
 asset, 83–84
 comparative advantage, 85–87
 free trade, 366
 macroeconomic measurement, 81–87
 trade deficit, 81–82
 trade surplus, 81–82
 traditional economy
 defined, 375
 described, 57
 long-standing cultural tradition, 57
 mixed economy, 56–57
 Tragedy of the Commons
 defined, 375
 environmental problems and poor
 property rights, 326–327
 overgrazing commonly owned field,
 324–326
 transfer payment, 78–79
 trusts, 306–307

• U •

Undergraduate Price Index, 101–102
 unemployment, 114
 uniformed decision-making, 34–35
 unintended consequence, 362
 upward price stickiness, 149
 upward wage stickiness, 149
 used car market
 asymmetric information, 331–335
 credit crunch, 331–335
 expert opinion, getting, 335
 poor quality, reasons for, 331–333
 reputation building, 334–335
 warranty, offering, 334
 utility
 cardinal, 199
 defined, 375
 ordinal, 199
 total, 33
 utility-maximising consumer. *See also*
 diminishing marginal utility
 buying as much as you can, 203–205
 cardinal versus ordinal utility, 199
 constrained optimisation, 198
 described, 197
 diminishing marginal utility, 199–202
 marginal utility, 32–33
 measuring utility, 27
 per pound marginal utility, equalising,
 208–211
 two goods, allocating money between,
 205–208

• V •

value added tax (VAT), 56
 variable cost, 224, 375

• W •

wage
 cutting wages versus cutting workers, 125
 downward wage stickiness, 149
 nominal, 147–148
 sticky prices, 126
 upward wage stickiness, 149
 Wal-Mart, 122
 Wall Street Crash, 128
 warranty, 334
The Wealth of Nations (Smith), 315
 welfare economics, 349
 Wendy's food production, 313
 work effort fallacy, 361
 worker. *See* labour
 WTO (World Trade Organisation), 84

• Z •

zero-coupon bond, 160



FOR DUMMIES[®]

Making Everything Easier!™

UK editions

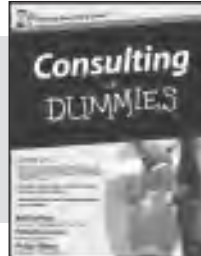
BUSINESS



978-0-470-74490-1



978-0-470-74381-2



978-0-470-71382-2

Anger Management For Dummies
978-0-470-68216-6

Boosting Self-Esteem For Dummies
978-0-470-74193-1

British Sign Language
For Dummies
978-0-470-69477-0

Business NLP For Dummies
978-0-470-69757-3

Cricket For Dummies
978-0-470-03454-5

CVs For Dummies, 2nd Edition
978-0-470-74491-8

Divorce For Dummies, 2nd Edition
978-0-470-74128-3

Emotional Freedom Technique
For Dummies
978-0-470-75876-2

Emotional Healing For Dummies
978-0-470-74764-3

English Grammar For Dummies
978-0-470-05752-0

Flirting For Dummies
978-0-470-74259-4

IBS For Dummies
978-0-470-51737-6

Improving Your Relationship For
Dummies
978-0-470-68472-6

Lean Six Sigma For Dummies
978-0-470-75626-3

Life Coaching For Dummies,
2nd Edition
978-0-470-66554-1

REFERENCE



978-0-470-68637-9



978-0-470-68837-3



978-0-470-74535-9

HOBBIES



978-0-470-69960-7



978-0-470-68641-6



978-0-470-68178-7

Available wherever books are sold. For more information or to order direct go to www.wiley.com or call +44 (0) 1243 843291

FOR DUMMIES®



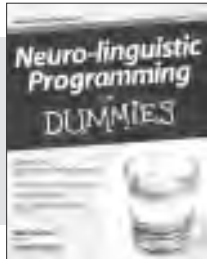
A world of resources to help you grow

UK editions

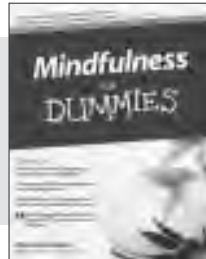
SELF-HELP



978-0-470-66541-1



978-0-470-66543-5



978-0-470-66086-7

Origami Kit For Dummies
978-0-470-75857-1

Overcoming Depression For Dummies
978-0-470-69430-5

Positive Psychology For Dummies
978-0-470-72136-0

PRINCE2 For Dummies, 2009 Edition
978-0-470-71025-8

Psychometric Tests For Dummies
978-0-470-75366-8

Raising Happy Children
For Dummies
978-0-470-05978-4

Reading the Financial Pages
For Dummies
978-0-470-71432-4

Sage 50 Accounts For Dummies
978-0-470-71558-1

Self-Hypnosis For Dummies
978-0-470-66073-7

Starting a Business For Dummies,
2nd Edition
978-0-470-51806-9

Study Skills For Dummies
978-0-470-74047-7

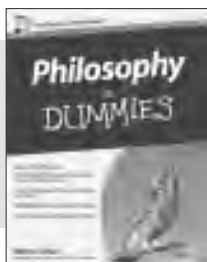
Teaching English as a Foreign
Language For Dummies
978-0-470-74576-2

Teaching Skills For Dummies
978-0-470-74084-2

Time Management For Dummies
978-0-470-77765-7

Work-Life Balance For Dummies
978-0-470-71380-8

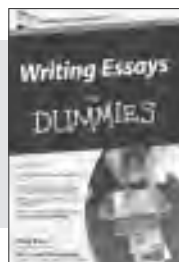
STUDENTS



978-0-470-68820-5



978-0-470-74711-7

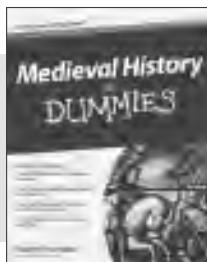


978-0-470-74290-7

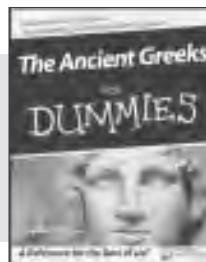
HISTORY



978-0-470-99468-9



978-0-470-74783-4



978-0-470-98787-2

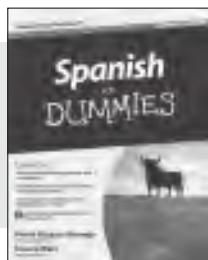
Available wherever books are sold. For more information or to order direct go to www.wiley.com
or call +44 (0) 1243 843291



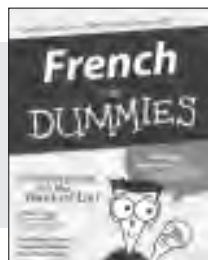
FOR DUMMIES[®]

The easy way to get more done and have more fun

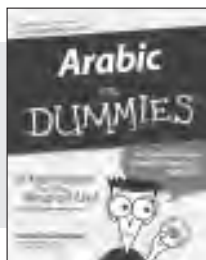
LANGUAGES



978-0-470-68815-1
UK Edition



978-0-7645-5193-2



978-0-471-77270-5

Art For Dummies
978-0-7645-5104-8

Bass Guitar For Dummies, 2nd Edition
978-0-470-53961-3

Christianity For Dummies
978-0-7645-4482-8

Criminology For Dummies
978-0-470-39696-4

Forensics For Dummies
978-0-7645-5580-0

German For Dummies
978-0-7645-5195-6

Hobby Farming For Dummies
978-0-470-28172-7

Index Investing For Dummies
978-0-470-29406-2

Knitting For Dummies, 2nd Edition
978-0-470-28747-7

Music Theory For Dummies
978-0-7645-7838-0

Piano For Dummies, 2nd Edition
978-0-470-49644-2

Physics For Dummies
978-0-7645-5433-9

Schizophrenia For Dummies
978-0-470-25927-6

Sex For Dummies, 3rd Edition
978-0-470-04523-7

Sherlock Holmes For Dummies
978-0-470-48444-9

Solar Power Your Home
For Dummies, 2nd Edition
978-0-470-59678-4

The Koran For Dummies
978-0-7645-5581-7

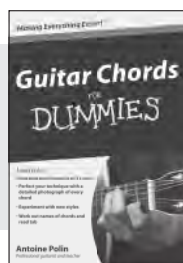
Wine All-in-One For Dummies
978-0-470-47626-0

Yoga For Dummies, 2nd Edition
978-0-470-50202-0

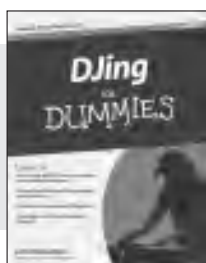
MUSIC



978-0-470-48133-2

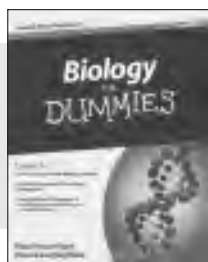


978-0-470-66603-6
Lay-flat, UK Edition

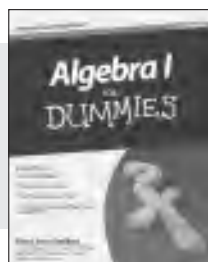


978-0-470-66372-1
UK Edition

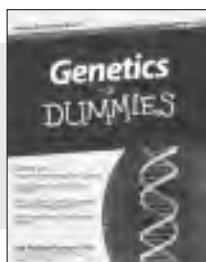
SCIENCE & MATHS



978-0-470-59875-7

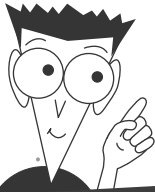


978-0-470-55964-2



978-0-470-55174-5

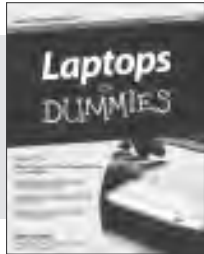
Available wherever books are sold. For more information or to order direct go to www.wiley.com or call +44 (0) 1243 843291



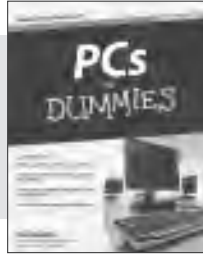
FOR DUMMIES[®]

Helping you expand your horizons and achieve your potential

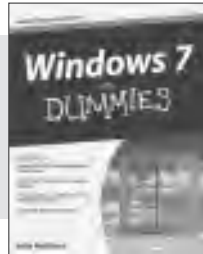
COMPUTER BASICS



978-0-470-57829-2



978-0-470-46542-4



978-0-470-49743-2

Access 2007 For Dummies
978-0-470-04612-8

Adobe Creative Suite 5 Design
Premium All-in-One For Dummies
978-0-470-60746-6

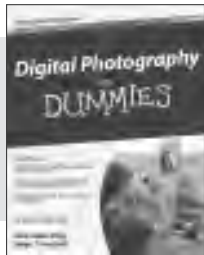
AutoCAD 2011 For Dummies
978-0-470-59539-8

C++ For Dummies, 6th Edition
978-0-470-31726-6

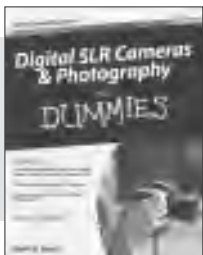
Computers For Seniors For Dummies,
2nd Edition
978-0-470-53483-0

Dreamweaver CS5 For Dummies
978-0-470-61076-3

DIGITAL PHOTOGRAPHY



978-0-470-25074-7



978-0-470-46606-3



978-0-470-59591-6

Excel 2007 All-In-One Desk Reference
For Dummies
978-0-470-03738-6

Green IT For Dummies
978-0-470-38688-0

Macs For Dummies, 10th Edition
978-0-470-27817-8

Mac OS X Snow Leopard For Dummies
978-0-470-43543-4

Networking All-in-One Desk Reference
For Dummies, 3rd Edition
978-0-470-17915-4

Photoshop CS5 For Dummies
978-0-470-61078-7

MICROSOFT OFFICE 2010



978-0-470-48998-7



978-0-470-58302-9



978-0-470-48953-6

Photoshop Elements 8 For Dummies
978-0-470-52967-6

Search Engine Optimization
For Dummies, 3rd Edition
978-0-470-26270-2

The Internet For Dummies,
12th Edition
978-0-470-56095-2

Visual Studio 2008 All-In-One Desk
Reference For Dummies
978-0-470-19108-8

Web Analytics For Dummies
978-0-470-09824-0

Available wherever books are sold. For more information or to order direct go to www.wiley.com
or call +44 (0) 1243 843291

FOR
DUMMIES[®]
Making Everything Easier![™]



**Get how-to
tips straight
to your inbox!**

Sign up to the monthly *For Dummies* e-newsletter to receive tips, special offers, competitions, sample chapters and more

www.dummies.com/go/UKnewsletter



Untangle the jargon and understand how you're involved in everyday economics

If you want to get to grips with the basics of economics and understand a subject that affects us all on a daily basis, then look no further than *Economics For Dummies*. This easy to understand guide takes you through the world of economics from understanding micro- and macroeconomics to demystifying complex topics such as capitalism and recession.

This updated edition walks you through the history, principles and theories of economics as well as breaking down all the complicated terminology, leaving you clued up on economics in no time.

- **Getting to grips** – explore the science of economics and how people deal with scarcity
- **Keeping an eye on it** – learn all about macroeconomics and how economists keep track of everything
- **Watch patterns emerge** – understand why monitoring consumer behaviour is vital and all you need to know about microeconomics
- **Your recession guide** – expert advice on recessions and a detailed look at why they occur

Peter Antonioni is a Senior Teaching Fellow in the Department of Management, Science and Innovation at University College London. He has worked in both the academic and private sectors as an economist.
Sean Masaki Flynn, PhD, is an Assistant Professor of Economics.



Open the book and find:

- Why you should care about economics and how it affects you
- Tools to help you understand a recession
- A guide to seductive economic fallacies
- All you need to know on monetary and fiscal policies
- How supply and demand can be made easy
- Why it's vital to track consumer choices
- An in-depth look at a profit-maximising firm and the core of capitalism
- Guidance on property rights and wrongs

Go to **Dummies.com**[®]
 for videos, step-by-step examples,
 how-to articles, or to shop!

For Dummies[®]
 A Branded Imprint of



£15.99 UK / \$26.99 US

ISBN 978-0-470-97325-7

