To Catherine, Nicholas, and Peter,
my other contributions to the next generation
About the Author

N. Gregory Mankiw is professor of economics at Harvard University. As a student, he studied economics at Princeton University and MIT. As a teacher, he has taught macroeconomics, microeconomics, statistics, and principles of economics. He even spent one summer long ago as a sailing instructor on Long Beach Island.

Professor Mankiw is a prolific writer and a regular participant in academic and policy debates. His work has been published in scholarly journals, such as the American Economic Review, Journal of Political Economy, and Quarterly Journal of Economics, and in more popular forums, such as The New York Times and The Wall Street Journal. He is also author of the best-selling intermediate-level textbook Macroeconomics (Worth Publishers). In addition to his teaching, research, and writing, Professor Mankiw has been a research associate of the National Bureau of Economic Research, an adviser to the Federal Reserve Bank of Boston and the Congressional Budget Office, and a member of the ETS test development committee for the Advanced Placement exam in economics. From 2003 to 2005, he served as chairman of the President’s Council of Economic Advisers.

Professor Mankiw lives in Wellesley, Massachusetts, with his wife, Deborah, three children, Catherine, Nicholas, and Peter, and their border terrier, Tobin.
**PART I  INTRODUCTION  1**  
**CHAPTER 1**  Ten Principles of Economics  3  
**CHAPTER 2**  Thinking Like an Economist  21  
**CHAPTER 3**  Interdependence and the Gains from Trade  49  

**PART II  HOW MARKETS WORK  63**  
**CHAPTER 4**  The Market Forces of Supply and Demand  65  
**CHAPTER 5**  Elasticity and Its Application  89  
**CHAPTER 6**  Supply, Demand, and Government Policies  113  

**PART III  MARKETS AND WELFARE  135**  
**CHAPTER 7**  Consumers, Producers, and the Efficiency of Markets  137  
**CHAPTER 8**  Application: The Costs of Taxation  159  
**CHAPTER 9**  Application: International Trade  177  

**PART IV  THE ECONOMICS OF THE PUBLIC SECTOR  201**  
**CHAPTER 10**  Externalities  203  
**CHAPTER 11**  Public Goods and Common Resources  225  

**PART V  FIRM BEHAVIOR AND THE ORGANIZATION OF INDUSTRY  241**  
**CHAPTER 12**  The Costs of Production  243  
**CHAPTER 13**  Firms in Competitive Markets  265  
**CHAPTER 14**  Monopoly  287  

**PART VI  THE DATA OF MACROECONOMICS  321**  
**CHAPTER 15**  Measuring a Nation’s Income  323  
**CHAPTER 16**  Measuring the Cost of Living  345  

**PART VII  THE REAL ECONOMY IN THE LONG RUN  363**  
**CHAPTER 17**  Production and Growth  365  
**CHAPTER 18**  Saving, Investment, and the Financial System  391  
**CHAPTER 19**  The Basic Tools of Finance  413  
**CHAPTER 20**  Unemployment  429  

**PART VIII  MONEY AND PRICES IN THE LONG RUN  455**  
**CHAPTER 21**  The Monetary System  457  
**CHAPTER 22**  Money Growth and Inflation  479  

**PART IX  SHORT-RUN ECONOMIC FLUCTUATIONS  505**  
**CHAPTER 23**  Aggregate Demand and Aggregate Supply  507  
**CHAPTER 24**  The Influence of Monetary and Fiscal Policy on Aggregate Demand  545
“Economics is a study of mankind in the ordinary business of life.” So wrote Alfred Marshall, the great 19th-century economist, in his textbook, *Principles of Economics*. Although we have learned much about the economy since Marshall’s time, this definition of economics is as true today as it was in 1890, when the first edition of his text was published.

Why should you, as a student at the beginning of the 21st century, embark on the study of economics? There are three reasons.

The first reason to study economics is that it will help you understand the world in which you live. There are many questions about the economy that might spark your curiosity. Why are apartments so hard to find in New York City? Why do airlines charge less for a round-trip ticket if the traveler stays over a Saturday night? Why is Johnny Depp paid so much to star in movies? Why are living standards so meager in many African countries? Why do some countries have high rates of inflation while others have stable prices? Why are jobs easy to find in some years and hard to find in others? These are just a few of the questions that a course in economics will help you answer.

The second reason to study economics is that it will make you a more astute participant in the economy. As you go about your life, you make many economic decisions. While you are a student, you decide how many years to stay in school. Once you take a job, you decide how much of your income to spend, how much to save, and how to invest your savings. Someday you may find yourself running a small business or a large corporation, and you will decide what prices to charge for your products. The insights developed in the coming chapters will give you a new perspective on how best to make these decisions. Studying economics will not by itself make you rich, but it will give you some tools that may help in that endeavor.

The third reason to study economics is that it will give you a better understanding of both the potential and the limits of economic policy. Economic questions are always on the minds of policymakers in mayors’ offices, governors’ mansions, and the White House. What are the burdens associated with alternative forms of taxation? What are the effects of free trade with other countries? What is the best way to protect the environment? How does a government budget deficit affect the economy? As a voter, you help choose the policies that guide the allocation of society’s resources. An understanding of economics will help you carry out that responsibility. And who knows: Perhaps someday you will end up as one of those policymakers yourself.

Thus, the principles of economics can be applied in many of life’s situations. Whether the future finds you reading the newspaper, running a business, or sitting in the Oval Office, you will be glad that you studied economics.

N. Gregory Mankiw
September 2008
We know you are often short on time. But you can maximize your efforts — and results — when you Experience Mankiw Fifth Edition’s engaging learning tools. With the product support website and EconCentral, you’ll quickly reinforce chapter concepts and sharpen your skills with interactive, hands-on applications online.

If a printed Study Guide better suits your needs and study habits, the Mankiw 5e Study Guide is unsurpassed in its careful attention to accuracy, concise language, and practice that enhances your study time.

**Product Support Website**

www.cengage.com/economics/mankiw

The Mankiw product support website is an easy online stop. Here, you’ll find access to free quizzing, Student Note Prompt handouts for the Premium PowerPoint® your instructor may use, and other resources.

**Study Guide**

Helping you achieve your personal best, the Mankiw Study Guide is based completely on the Fifth Edition, covering chapter material comprehensively — and accurately. Very hands-on, each chapter thoroughly covers the material in the corresponding chapter of Mankiw. Every key word and concept is addressed within the Study Guide chapter — meaning you’ll feel confident that if you can do the study guide, you will understand all of the material in that chapter of Mankiw.

The “types” of questions used in the Study Guide reflect what you find most useful when studying. Our student surveys show that students like you felt that fill-in-the-blank questions, matching questions, and questions without specific single answers were an inefficient use of their time — and the Mankiw Study Guide avoids these kinds of questions.

To order the study guide, visit www.ichapters.com.
Multiple resources for learning and reinforcing principles concepts are now available in one place! EconCentral is your one-stop shop for the learning tools and activities to help you succeed.

At a minimal extra cost, EconCentral equips you with a portal to a wealth of resources that help you both study and apply economic concepts. As you read and study the chapters, you can access video tutorials with Greg Mankiw Answers Key Questions, 10 Principles Videos, and Ask the Instructor Videos. You can review with Flash Cards and the Graphing Workshop, check your understanding of the chapter with interactive quizzing, and print Student Note Prompt handouts for the Premium PowerPoint® to make note-taking in class much more efficient.

Ready to apply chapter concepts to the real world? EconCentral gives you ABC News videos, EconNews articles, Economic debates, Links to Economic Data, and more. All the study and application resources in EconCentral are organized by chapter to help you get the most from the Mankiw text and from your classes.

Visit www.cengage.com/economics/mankiw/5e/econcentral to see the study options available!
In writing this book, I benefited from the input of many talented people. Indeed, the list of people who have contributed to this project is so long, and their contributions so valuable, that it seems an injustice that only a single name appears on the cover.

Let me begin with my colleagues in the economics profession. The four editions of this text and its supplemental materials have benefited enormously from their input. In reviews and surveys, they have offered suggestions, identified challenges, and shared ideas from their own classroom experience. I am indebted to them for the perspectives they have brought to the text. Unfortunately, the list has become too long to thank those who contributed to previous editions, even though students reading the current edition are still benefiting from their insights.

Most important in this process have been Ron Cronovich (Carthage College) and David Hakes (University of Northern Iowa). Ron and David, both dedicated teachers, have served as reliable sounding boards for ideas and hardworking partners with me in putting together the superb package of supplements.

For this new edition, the following diary reviewers recorded their day-to-day experience over the course of a semester, offering detailed suggestions about how to improve the text.

John Crooker, University of Central Missouri
Rachel Friedberg, Brown University
Greg Hunter, California State University, Polytechnic, Pomona
Lillian Kamal, Northwestern University
Francis Kemegue, Bryant University
Douglas Miller, University of Missouri
Babu Nahata, University of Louisville
Edward Skelton, Southern Methodist University

The following reviewers of the fourth edition provided suggestions for refining the content, organization, and approach in the fifth.

Syed Ahmed, Cameron University
Farhad Ameen, State University of New York, Westchester Community College
Mohammad Bajwa, Northampton Community College
Carl Bauer, Oakton Community College
Roberta Bibly, Grand Valley State University
Stephen Billings, University of Colorado at Boulder
Bruce Brown, California State University, Polytechnic, Pomona
Lynn Burbridge, Northern Kentucky University
Mark Chester, Reading Area Community College
David Ching, University of Hawaii, Manoa
Sarah Cosgrove, University of Massachusetts, Dartmouth
I received detailed feedback on specific elements in the text, including all end-of-chapter problems and applications, from the following instructors.

Casey R. Abington, Kansas State University
Seemi Ahmad, Dutchess Community College
Farhad Ameen, State University of New York, Westchester Community College
J. J. Arias, Georgia College & State University
James Bathgate, Willamette University
Scott Beaulier, Mercer University
Clive Belfield, Queens College
Calvin Blackwell, College of Charleston
Cecil E. Bohanon, Ball State University
Douglas Campbell, University of Memphis
Michael G. Carew, Baruch College
Sewin Chan, New York University
Joyce J. Chen, The Ohio State University
Edward A. Cohn, Del Mar College
Chad D. Cotti, University of South Carolina
Erik D. Craft, University of Richmond
Eleanor D. Craig, University of Delaware
Abdelmagead Elbiali, Rio Hondo College
Harold W. Elder, University of Alabama
Hadi Salehi Esfahani, University of Illinois, Urbana-Champaign
David Franck, Francis Marion University
Amanda S. Freeman, Kansas State University
J.P. Gilbert, Miracosta College
Joanne Guo, Dyson College of Pace University
Charles E. Hegji, Auburn University at Montgomery
Andrew J. Hussey, University of Memphis
Hans R. Isakson, University of Northern Iowa
Simran Kahai, John Carroll University
David E. Kalist, Shippensburg University
Mark P. Karscig, University of Central Missouri
Theodore Kuhn, Butler University
Dong Li, Kansas State University
Daniel Lin, George Mason University
Nathaniel Manning, Southern University
Vince Marra, University of Delaware
Akbar Marvasti, University of Southern Mississippi
Heather Mattson, University of Saint Thomas
Charles C. Moul, Washington University in St. Louis
Albert A. Okunade, University of Memphis
J. Brian O’Roark, Robert Morris University
Anthony L. Ostrosky, Illinois State University
Nitin V. Paranjpe, Wayne State University & Oakland University
Sanela Porca, University of South Carolina, Aiken
Walter G. Park, American University
Reza M. Ramazani, Saint Michael’s College
Rhonda Vonshay Sharpe, University of Vermont
Carolyn Fabian Stumph, Indiana University–Purdue University Fort Wayne
Rick Tannery, Slippery Rock University
Aditi Thapar, New York University
Michael H. Tew, Troy University
Jennifer A. Vincent, Champlain College
Milos Vulanovic, Lehman College
Bhavneet Walia, Kansas State University
Douglas M. Walker, College of Charleston
Patrick Walsh, Saint Michael’s College
Larry Wolfenbarger, Macon State College
William C. Wood, James Madison University
Chiou-nan Yeh, Alabama State University

The accuracy of a textbook is critically important. I am responsible for any remaining errors, but I am grateful to the following professors for reading through the final manuscript and page proofs with me:

Joel Dalafave, Bucks County Community College
Greg Hunter, California State University – Pomona
Lillian Kamal, Northwestern University

Francis Kemegue, Bryant University
Douglas Miller, University of Missouri
Ed Skelton, Southern Methodist University

The team of editors who worked on this book improved it tremendously. Jane Tufts, developmental editor, provided truly spectacular editing—as she always does. Mike Worls, economics executive editor, did a splendid job of overseeing the many people involved in such a large project. Jennifer Thomas (senior developmental editor) and Katie Yanos (developmental editor) were crucial in assembling an extensive and thoughtful group of reviewers to give me feedback on the
previous edition, while putting together an excellent team to revise the supplements. Colleen Farmer, senior content project manager, and Katherine Wilson, senior project manager, had the patience and dedication necessary to turn my manuscript into this book. Michelle Kunkler, senior art director, gave this book its clean, friendly look. Michael Steirnagle, the illustrator, helped make the book more visually appealing and the economics in it less abstract. Carolyn Crabtree, copyeditor, refined my prose, and Terry Casey, indexer, prepared a careful and thorough index. Brian Joyner, executive marketing manager, worked long hours getting the word out to potential users of this book. The rest of the Cengage team, including Jean Butstrom, Sandra Milewski and Deepak Kumar, was also consistently professional, enthusiastic, and dedicated.

I am grateful also to Josh Bookin, a former Advanced Placement economics teacher and recently an extraordinary section leader for Ec 10, the introductory course at Harvard. Josh helped me refine the manuscript and check the page proofs for this edition.

As always, I must thank my “in-house” editor Deborah Mankiw. As the first reader of almost everything I write, she continued to offer just the right mix of criticism and encouragement.

Finally, I would like to mention my three children Catherine, Nicholas, and Peter. Their contribution to this book was putting up with a father spending too many hours in his study. The four of us have much in common—not least of which is our love of ice cream (which becomes apparent in Chapter 4). Maybe sometime soon one of them will pick up my passion for economics as well.

N. Gregory Mankiw
September 2008
# Table of Contents

Preface: To the Student vii

## PART I

### INTRODUCTION 1

### CHAPTER 1

#### TEN PRINCIPLES OF ECONOMICS 3

**How People Make Decisions** 4
- Principle 1: People Face Trade-offs 4
- Principle 2: The Cost of Something Is What You Give Up to Get It 5
- Principle 3: Rational People Think at the Margin 6
- Principle 4: People Respond to Incentives 7

**How People Interact** 8
- Principle 5: Trade Can Make Everyone Better Off 8
- Principle 6: Markets Are Usually a Good Way to Organize Economic Activity 8
- **IN THE NEWS** Incentive Pay 9
- Principle 7: Governments Can Sometimes Improve Market Outcomes 10
- **FYI** Adam Smith and the Invisible Hand 11

**How the Economy as a Whole Works** 12
- Principle 8: A Country’s Standard of Living Depends on Its Ability to Produce Goods and Services 12
- Principle 9: Prices Rise When the Government Prints Too Much Money 13
- **IN THE NEWS** Why You Should Study Economics 14
- Principle 10: Society Faces a Short-Run Trade-off between Inflation and Unemployment 14

**Conclusion** 15
- **FYI** How to Read This Book 16

### Summary 17

### Key Concepts 17

### Questions for Review 18

### Problems and Applications 18

## CHAPTER 2

### THINKING LIKE AN ECONOMIST 21

**The Economist as Scientist** 22
- The Scientific Method: Observation, Theory, and More Observation 22
- The Role of Assumptions 23
- Economic Models 23
- Our First Model: The Circular-Flow Diagram 24
- Our Second Model: The Production Possibilities Frontier 25
- Microeconomics and Macroeconomics 28
- **FYI** Who Studies Economics? 29

**The Economist as Policy Adviser** 30
- Positive versus Normative Analysis 30
- Economists in Washington 31
- **IN THE NEWS** Football Economics 32
- Why Economists’ Advice Is Not Always Followed 32

**Why Economists Disagree** 34
<table>
<thead>
<tr>
<th>Chapter 8</th>
<th>APPLICATION: THE COSTS OF TAXATION</th>
<th>159</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Deadweight Loss of Taxation</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>How a Tax Affects Market Participants</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>Deadweight Losses and the Gains from Trade</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>The Determinants of the Deadweight Loss</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>CASE STUDY The Deadweight Loss Debate</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td>Deadweight Loss and Tax Revenue as Taxes Vary</td>
<td>167</td>
<td></td>
</tr>
<tr>
<td>FYI Henry George and the Land Tax</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>CASE STUDY The Laffer Curve and Supply-Side Economics</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>IN THE NEWS On the Way to France</td>
<td>170</td>
<td></td>
</tr>
</tbody>
</table>

Conclusion 172
Summary 172

| Key Concepts 173
Questions for Review 173
Problems and Applications 173

CHAPTER 9
APPLICATION: INTERNATIONAL TRADE 177

| The Determinants of Trade | 178 |
| The Equilibrium without Trade | 178 |
| The World Price and Comparative Advantage | 179 |
| The Winners and Losers from Trade | 180 |
| The Gains and Losses of an Exporting Country | 180 |
| The Gains and Losses of an Importing Country | 181 |
| The Effects of a Tariff | 183 |
| The Lessons for Trade Policy | 185 |
| FYI Import Quotas: Another Way to Restrict Trade | 185 |
| Other Benefits of International Trade | 186 |
| IN THE NEWS Should the Winners from Free Trade Compensate the Losers? | 187 |
| The Arguments for Restricting Trade | 188 |
| The Jobs Argument | 188 |
| IN THE NEWS Offshore Outsourcing | 189 |
| The National-Security Argument | 190 |
| The Infant-Industry Argument | 190 |
| The Unfair-Competition Argument | 191 |
| The Protection-as-a-Bargaining-Chip Argument | 191 |
| IN THE NEWS Second Thoughts about Free Trade | 192 |
| CASE STUDY Trade Agreements and the World Trade Organization | 192 |

Conclusion 194
Summary 195
Key Concepts 196
Questions for Review 196
Problems and Applications 196
PART IV
THE ECONOMICS OF THE PUBLIC SECTOR  201

CHAPTER 10
EXTERNALITIES  203

Externalities and Market Inefficiency  204
  Welfare Economics: A Recap  205
  Negative Externalities  205
  Positive Externalities  207
  CASE STUDY Technology Spillovers, Industrial Policy, and Patent Protection  208

Public Policies toward Externalities  209
  Command-and-Control Policies: Regulation  209
  Market-Based Policy 1: Corrective Taxes and Subsidies  210
  CASE STUDY Why Is Gasoline Taxed So Heavily?  211
  Market-Based Policy 2: Tradable Pollution Permits  212
  Objections to the Economic Analysis of Pollution  214

Private Solutions to Externalities  215
  The Types of Private Solutions  215
  IN THE NEWS The Case for Taxing Carbon  216
  The Coase Theorem  217
  Why Private Solutions Do Not Always Work  218

Conclusion  219
Summary  220
Key Concepts  221
Questions for Review  221
Problems and Applications  221

CHAPTER 11
PUBLIC GOODS AND COMMON RESOURCES  225

The Different Kinds of Goods  226

Public Goods  227
  The Free-Rider Problem  228
  Some Important Public Goods  228
  CASE STUDY Are Lighthouses Public Goods?  230
  The Difficult Job of Cost–Benefit Analysis  230
  CASE STUDY How Much Is a Life Worth?  231

Common Resources  232
  The Tragedy of the Commons  232
  Some Important Common Resources  233
  IN THE NEWS The Bloomberg Plan  234
  CASE STUDY Why the Cow Is Not Extinct  236

Conclusion: The Importance of Property Rights  237
Summary  238
Key Concepts  238
Questions for Review  238
Problems and Applications  238

PART V
FIRM BEHAVIOR AND THE ORGANIZATION OF INDUSTRY  241

CHAPTER 12
THE COSTS OF PRODUCTION  243

What Are Costs?  244
  Total Revenue, Total Cost, and Profit  244
Costs as Opportunity Costs 244
The Cost of Capital as an Opportunity Cost 245
Economic Profit versus Accounting Profit 246

Production and Costs 247
The Production Function 247
From the Production Function to the Total-Cost Curve 249

The Various Measures of Cost 250
Fixed and Variable Costs 250
Average and Marginal Cost 251
Cost Curves and Their Shapes 252
Typical Cost Curves 254

Costs in the Short Run and in the Long Run 256
The Relationship between Short-Run and Long-Run Average Total Cost 256
Economies and Diseconomies of Scale 257
FYI Lessons from a Pin Factory 258

Conclusion 258
Summary 259
Key Concepts 260
Questions for Review 260
Problems and Applications 261

CHAPTER 13
FIRMS IN COMPETITIVE MARKETS 265

What Is a Competitive Market? 266
The Meaning of Competition 266
The Revenue of a Competitive Firm 266

Profit Maximization and the Competitive Firm’s Supply Curve 268
A Simple Example of Profit Maximization 268
The Marginal-Cost Curve and the Firm’s Supply Decision 269
The Firm’s Short-Run Decision to Shut Down 271
Spilt Milk and Other Sunk Costs 272
CASE STUDY Near-Empty Restaurants and Off-Season Miniature Golf 273
The Firm’s Long-Run Decision to Exit or Enter a Market 274

Measuring Profit in Our Graph for the Competitive Firm 275

The Supply Curve in a Competitive Market 277
The Short Run: Market Supply with a Fixed Number of Firms 277
The Long Run: Market Supply with Entry and Exit 278
Why Do Competitive Firms Stay in Business If They Make Zero Profit? 279
A Shift in Demand in the Short Run and Long Run 280
Why the Long-Run Supply Curve Might Slope Upward 280

Conclusion: Behind the Supply Curve 282
Summary 283
Key Concepts 283
Questions for Review 283
Problems and Applications 284

CHAPTER 14
MONOPOLY 287

Why Monopolies Arise 288
Monopoly Resources 289
Government-Created Monopolies 289
Natural Monopolies 290

How Monopolies Make Production and Pricing Decisions 291
Monopoly versus Competition 291
A Monopoly’s Revenue 292
Profit Maximization 295
FYI Why a Monopoly Does Not Have a Supply Curve 296
A Monopoly’s Profit 296
CASE STUDY Monopoly Drugs versus Generic Drugs 297

The Welfare Cost of Monopolies 298
The Deadweight Loss 299
The Monopoly’s Profit: A Social Cost? 301

Price Discrimination 302
A Parable about Pricing 302
The Moral of the Story 303
The Analytics of Price Discrimination 304
Examples of Price Discrimination 305
IN THE NEWS TKTS and Other Schemes 306

Public Policy toward Monopolies 308
Increasing Competition with Antitrust Laws 308
Regulation 309
IN THE NEWS Airline Mergers 309
IN THE NEWS Public Transport and Private Enterprise 310
Public Ownership 312
Doing Nothing 313

Conclusion: The Prevalence of Monopolies 314
Summary 314
Key Concepts 315
Questions for Review 315
Problems and Applications 316

PART VI
THE DATA OF MACROECONOMICS 321

CHAPTER 15
MEASURING A NATION’S INCOME 323

The Economy’s Income and Expenditure 324

The Measurement of Gross Domestic Product 326
“GDP Is the Market Value…” 326
“. . . of All . . .” 326
“. . . Final . . .” 327
“. . . Goods and Services . . .” 327
“. . . Produced . . .” 327
“. . . Within a Country . . .” 327
“. . . In a Given Period of Time.” 327

The Components of GDP 328
Consumption 329
Investment 329
FYI Other Measures of Income 329
Government Purchases 330
Net Exports 330
CASE STUDY The Components of U.S. GDP 331

Real versus Nominal GDP 331
A Numerical Example 332
The GDP Deflator 333
CASE STUDY Real GDP Over Recent History 334

Is GDP a Good Measure of Economic Well-Being? 335
IN THE NEWS The Underground Economy 336
CASE STUDY International Differences in GDP and the Quality of Life 338

Conclusion 339
FYI Who Wins at the Olympics? 340

Summary 340
Key Concepts 341
Questions for Review 341
Problems and Applications 341

CHAPTER 16
MEASURING THE COST OF LIVING 345

The Consumer Price Index 346
How the Consumer Price Index Is Calculated 346
Problems in Measuring the Cost of Living 348
FYI What Is In the CPI’s Basket? 349
IN THE NEWS Accounting for Quality Change 350
The GDP Deflator versus the Consumer Price Index 352

Correcting Economic Variables for the Effects of Inflation 353
Dollar Figures from Different Times 354
Indexation 354
Real and Nominal Interest Rates 355
FYI Mr. Index Goes to Hollywood 355
PART VII
THE REAL ECONOMY IN THE LONG RUN 363

CHAPTER 17
PRODUCTION AND GROWTH 365

Economic Growth around the World 366
   FYI A Picture Is Worth a Thousand Statistics 368
   FYI Are You Richer Than the Richest American? 370

Productivity: Its Role and Determinants 370
   Why Productivity Is So Important 370
   How Productivity Is Determined 371
   FYI The Production Function 373
   IN THE NEWS Measuring Capital 374
   CASE STUDY Are Natural Resources a Limit to Growth? 374

Economic Growth and Public Policy 376
   Saving and Investment 376
   Diminishing Returns and the Catch-Up Effect 377
   Investment from Abroad 378
   Education 379

IN THE NEWS Promoting Human Capital 380
   Health and Nutrition 381
   Property Rights and Political Stability 382
   Free Trade 383
   Research and Development 384
   Population Growth 384
   IN THE NEWS Escape from Malthus 386

Conclusion: The Importance of Long-Run Growth 388
   Summary 389
   Key Concepts 389
   Questions for Review 389
   Problems and Applications 390

CHAPTER 18
SAVING, INVESTMENT, AND THE FINANCIAL SYSTEM 391

Financial Institutions in the U.S. Economy 392
   Financial Markets 392
   Financial Intermediaries 394
   FYI Key Numbers for Stock Watchers 395
   Summing Up 396

Saving and Investment in the National Income Accounts 397
   Some Important Identities 397
   The Meaning of Saving and Investment 399

The Market for Loanable Funds 399
   Supply and Demand for Loanable Funds 400
   Policy 1: Saving Incentives 402
   Policy 2: Investment Incentives 403
   IN THE NEWS In Praise of Misers 404
   Policy 3: Government Budget Deficits and Surpluses 405
   CASE STUDY The History of U.S. Government Debt 407

Conclusion 409
   Summary 410
   Key Concepts 410
   Questions for Review 410
   Problems and Applications 411
CHAPTER 19
THE BASIC TOOLS OF FINANCE  413

Present Value: Measuring the Time Value of Money  414
  FYI The Magic of Compounding and the Rule
  of 70  416

Managing Risk  416
  Risk Aversion  417
  The Markets for Insurance  417
  FYI The Peculiarities of Health Insurance  418
  Diversification of Firm-Specific Risk  419
  The Trade-off between Risk and Return  420

Asset Valuation  422
  Fundamental Analysis  422
  CASE STUDY Random Walks and Index Funds  423
  The Efficient Markets Hypothesis  422
  IN THE NEWS Neurofinance  424
  Market Irrationality  425

Conclusion  426
Summary  427
Key Concepts  427
Questions for Review  427
Problems and Applications  427

CHAPTER 20
UNEMPLOYMENT  429

Identifying Unemployment  430
  How Is Unemployment Measured?  430
  CASE STUDY Labor-Force Participation of Men and
  Women in the U.S. Economy  433
  IN THE NEWS The Rise of Adult Male
  Joblessness  434
  Does the Unemployment Rate Measure What We
  Want It To?  435
  How Long Are the Unemployed without Work?  437
  Why Are There Always Some People Unemployed?  438

Job Search  439
  Why Some Frictional Unemployment Is Inevitable  439
  FYI The Jobs Number  439
  Public Policy and Job Search  440

PART VIII
MONEY AND PRICES IN
THE LONG RUN  455

CHAPTER 21
THE MONETARY SYSTEM  457

The Meaning of Money  458
  The Functions of Money  458
The Reality of Short-Run Fluctuations 512
The Model of Aggregate Demand and Aggregate Supply 513

The Aggregate-Demand Curve 514
Why the Aggregate-Demand Curve Slopes Downward 514
Why the Aggregate-Demand Curve Might Shift 517
IN THE NEWS The 2008 Fiscal Stimulus 518

The Aggregate-Supply Curve 519
Why the Aggregate-Supply Curve Is Vertical in the Long Run 520
Why the Long-Run Aggregate-Supply Curve Might Shift 521
Using Aggregate Demand and Aggregate Supply to Depict Long-Run Growth and Inflation 523
Why the Aggregate-Supply Curve Slopes Upward in the Short Run 523
Why the Short-Run Aggregate-Supply Curve Might Shift 528

Two Causes of Economic Fluctuations 529
The Effects of a Shift in Aggregate Demand 530
FYI Monetary Neutrality Revisited 532
CASE STUDY Two Big Shifts in Aggregate Demand: The Great Depression and World War II 533
CASE STUDY The Recession of 2001 535
The Effects of a Shift in Aggregate Supply 535
FYI The Origins of Aggregate Demand and Aggregate Supply 538
CASE STUDY Oil and the Economy 538

Conclusion 539
Summary 540
Key Concepts 541
Questions for Review 541
Problems and Applications 541

CHAPTER 24
THE INFLUENCE OF MONETARY AND FISCAL POLICY ON AGGREGATE DEMAND 545

How Monetary Policy Influences Aggregate Demand 546
The Theory of Liquidity Preference 547
The Downward Slope of the Aggregate-Demand Curve 549
FYI Interest Rates in the Long Run and the Short Run 550
Changes in the Money Supply 551
The Role of Interest-Rate Targets in Fed Policy 553
IN THE NEWS The FOMC Explains Itself 554
CASE STUDY Why the Fed Watches the Stock Market (and Vice Versa) 554

How Fiscal Policy Influences Aggregate Demand 555
Changes in Government Purchases 555
The Multiplier Effect 556
A Formula for the Spending Multiplier 556
Other Applications of the Multiplier Effect 558
The Crowding-Out Effect 559
Changes in Taxes 560

Using Policy to Stabilize the Economy 561
FYI How Fiscal Policy Might Affect Aggregate Supply 561
The Case for Active Stabilization Policy 562
CASE STUDY Keynesians in the White House 563
The Case against Active Stabilization Policy 563
Automatic Stabilizers 565

Conclusion 565
Summary 566
Key Concepts 566
Questions for Review 567
Problems and Applications 567

Glossary 569
Index 574
PART I
Introduction
This page intentionally left blank
The word *economy* comes from the Greek word *oikonomos*, which means “one who manages a household.” At first, this origin might seem peculiar. But in fact, households and economies have much in common.

A household faces many decisions. It must decide which members of the household do which tasks and what each member gets in return: Who cooks dinner? Who does the laundry? Who gets the extra dessert at dinner? Who gets to choose what TV show to watch? In short, the household must allocate its scarce resources among its various members, taking into account each member’s abilities, efforts, and desires.

Like a household, a society faces many decisions. A society must find some way to decide what jobs will be done and who will do them. It needs some people to grow food, other people to make clothing, and still others to design computer software. Once society has allocated people (as well as land, buildings, and machines) to various jobs, it must also allocate the output of goods and services they produce. It must decide who will eat caviar and who will eat potatoes. It must decide who will drive a Ferrari and who will take the bus.

The management of society’s resources is important because resources are scarce. **Scarcity** means that society has limited resources and therefore cannot produce all the goods and services people wish to have. Just as each member of a household cannot get everything he or she wants, each individual in a society cannot attain the highest standard of living to which he or she might aspire.
Economics is the study of how society manages its scarce resources. In most societies, resources are allocated not by an all-powerful dictator but through the combined actions of millions of households and firms. Economists therefore study how people make decisions: how much they work, what they buy, how much they save, and how they invest their savings. Economists also study how people interact with one another. For instance, they examine how the multitude of buyers and sellers of a good together determine the price at which the good is sold and the quantity that is sold. Finally, economists analyze forces and trends that affect the economy as a whole, including the growth in average income, the fraction of the population that cannot find work, and the rate at which prices are rising.

The study of economics has many facets, but it is unified by several central ideas. In this chapter, we look at Ten Principles of Economics. Don’t worry if you don’t understand them all at first or if you aren’t completely convinced. We will explore these ideas more fully in later chapters. The ten principles are introduced here to give you an overview of what economics is all about. Consider this chapter a “preview of coming attractions.”

**HOW PEOPLE MAKE DECISIONS**

There is no mystery to what an economy is. Whether we are talking about the economy of Los Angeles, the United States, or the whole world, an economy is just a group of people dealing with one another as they go about their lives. Because the behavior of an economy reflects the behavior of the individuals who make up the economy, we begin our study of economics with four principles of individual decision making.

**PRINCIPLE 1: PEOPLE FACE TRADE-OFFS**

You may have heard the old saying, “There ain’t no such thing as a free lunch.” Grammar aside, there is much truth to this adage. To get one thing that we like, we usually have to give up another thing that we like. Making decisions requires trading off one goal against another.

Consider a student who must decide how to allocate her most valuable resource—her time. She can spend all her time studying economics, spend all of it studying psychology, or divide it between the two fields. For every hour she studies one subject, she gives up an hour she could have used studying the other. And for every hour she spends studying, she gives up an hour that she could have spent napping, bike riding, watching TV, or working at her part-time job for some extra spending money.

Or consider parents deciding how to spend their family income. They can buy food, clothing, or a family vacation. Or they can save some of the family income for retirement or the children’s college education. When they choose to spend an extra dollar on one of these goods, they have one less dollar to spend on some other good.

When people are grouped into societies, they face different kinds of trade-offs. The classic trade-off is between “guns and butter.” The more a society spends on national defense (guns) to protect its shores from foreign aggressors, the less it can spend on consumer goods (butter) to raise the standard of living at home. Also important in modern society is the trade-off between a clean environment and a high level of income. Laws that require firms to reduce pollution raise the
cost of producing goods and services. Because of the higher costs, these firms end up earning smaller profits, paying lower wages, charging higher prices, or some combination of these three. Thus, while pollution regulations yield the benefit of a cleaner environment and the improved health that comes with it, they have the cost of reducing the incomes of the firms’ owners, workers, and customers.

Another trade-off society faces is between efficiency and equality. Efficiency means that society is getting the maximum benefits from its scarce resources. Equality means that those benefits are distributed uniformly among society’s members. In other words, efficiency refers to the size of the economic pie, and equality refers to how the pie is divided into individual slices.

When government policies are designed, these two goals often conflict. Consider, for instance, policies aimed at equalizing the distribution of economic well-being. Some of these policies, such as the welfare system or unemployment insurance, try to help the members of society who are most in need. Others, such as the individual income tax, ask the financially successful to contribute more than others to support the government. While achieving greater equality, these policies reduce efficiency. When the government redistributes income from the rich to the poor, it reduces the reward for working hard; as a result, people work less and produce fewer goods and services. In other words, when the government tries to cut the economic pie into more equal slices, the pie gets smaller.

Recognizing that people face trade-offs does not by itself tell us what decisions they will or should make. A student should not abandon the study of psychology just because doing so would increase the time available for the study of economics. Society should not stop protecting the environment just because environmental regulations reduce our material standard of living. The poor should not be ignored just because helping them distorts work incentives. Nonetheless, people are likely to make good decisions only if they understand the options they have available. Our study of economics, therefore, starts by acknowledging life’s trade-offs.

**Principle 2: The Cost of Something Is What You Give Up to Get It**

Because people face trade-offs, making decisions requires comparing the costs and benefits of alternative courses of action. In many cases, however, the cost of an action is not as obvious as it might first appear.

Consider the decision to go to college. The main benefits are intellectual enrichment and a lifetime of better job opportunities. But what are the costs? To answer this question, you might be tempted to add up the money you spend on tuition, books, room, and board. Yet this total does not truly represent what you give up to spend a year in college.

There are two problems with this calculation. First, it includes some things that are not really costs of going to college. Even if you quit school, you need a place to sleep and food to eat. Room and board are costs of going to college only to the extent that they are more expensive at college than elsewhere. Second, this calculation ignores the largest cost of going to college—your time. When you spend a year listening to lectures, reading textbooks, and writing papers, you cannot spend that time working at a job. For most students, the earnings given up to attend school are the largest single cost of their education.

The opportunity cost of an item is what you give up to get that item. When making any decision, decision makers should be aware of the opportunity costs
that accompany each possible action. In fact, they usually are. College athletes
who can earn millions if they drop out of school and play professional sports are
well aware that their opportunity cost of college is very high. It is not surprising
that they often decide that the benefit is not worth the cost.

**PRINCIPLE 3: RATIONAL PEOPLE THINK AT THE MARGIN**

Economists normally assume that people are rational. **Rational people** systematically
and purposefully do the best they can to achieve their objectives, given the
available opportunities. As you study economics, you will encounter firms that
decide how many workers to hire and how much of their product to manufacture
and sell to maximize profits. You will also encounter individuals who decide how
much time to spend working and what goods and services to buy with the resulting
income to achieve the highest possible level of satisfaction.

Rational people know that decisions in life are rarely black and white but usu-
ally involve shades of gray. At dinnertime, the decision you face is not between
fasting or eating like a pig but whether to take that extra spoonful of mashed po-
tatoes. When exams roll around, your decision is not between blowing them off or
studying 24 hours a day but whether to spend an extra hour reviewing your notes
instead of watching TV. Economists use the term **marginal changes** to describe
small incremental adjustments to an existing plan of action. Keep in mind that
**margin** means “edge,” so marginal changes are adjustments around the edges of
what you are doing. Rational people often make decisions by comparing **marginal
benefits** and **marginal costs**.

For example, consider an airline deciding how much to charge passengers who
fly standby. Suppose that flying a 200-seat plane across the United States costs the
airline $100,000. In this case, the average cost of each seat is $100,000/200, which is
$500. One might be tempted to conclude that the airline should never sell a ticket
for less than $500. In fact, a rational airline can often find ways to raise its profits
by thinking at the margin. Imagine that a plane is about to take off with ten empty
seats, and a standby passenger waiting at the gate will pay $300 for a seat. Should
the airline sell the ticket? Of course it should. If the plane has empty seats, the cost
of adding one more passenger is tiny. Although the **average** cost of flying a pas-
enger is $500, the **marginal** cost is merely the cost of the bag of peanuts and can
of soda that the extra passenger will consume. As long as the standby passenger
pays more than the marginal cost, selling the ticket is profitable.

Marginal decision making can help explain some otherwise puzzling economic
phenomena. Here is a classic question: Why is water so cheap, while diamonds
are so expensive? Humans need water to survive, while diamonds are unneces-
sary; but for some reason, people are willing to pay much more for a diamond
than for a cup of water. The reason is that a person’s willingness to pay for any
good is based on the marginal benefit that an extra unit of the good would yield.
The marginal benefit, in turn, depends on how many units a person already has.
Water is essential, but the marginal benefit of an extra cup is small because water
is plentiful. By contrast, no one needs diamonds to survive, but because diamonds
are so rare, people consider the marginal benefit of an extra diamond to be large.

A rational decision maker takes an action if and only if the marginal benefit of the
action exceeds the marginal cost. This principle can explain why airlines are will-
ing to sell a ticket below average cost and why people are willing to pay more for
diamonds than for water. It can take some time to get used to the logic of marginal
thinking, but the study of economics will give you ample opportunity to practice.
PRINCIPLE 4: PEOPLE RESPOND TO INCENTIVES

An incentive is something that induces a person to act, such as the prospect of a punishment or a reward. Because rational people make decisions by comparing costs and benefits, they respond to incentives. You will see that incentives play a central role in the study of economics. One economist went so far as to suggest that the entire field could be simply summarized: “People respond to incentives. The rest is commentary.”

Incentives are crucial to analyzing how markets work. For example, when the price of an apple rises, people decide to eat fewer apples. At the same time, apple orchards decide to hire more workers and harvest more apples. In other words, a higher price in a market provides an incentive for buyers to consume less and an incentive for sellers to produce more. As we will see, the influence of prices on the behavior of consumers and producers is crucial for how a market economy allocates scarce resources.

Public policymakers should never forget about incentives: Many policies change the costs or benefits that people face and, therefore, alter their behavior. A tax on gasoline, for instance, encourages people to drive smaller, more fuel-efficient cars. That is one reason people drive smaller cars in Europe, where gasoline taxes are high, than in the United States, where gasoline taxes are low. A gasoline tax also encourages people to carpool, take public transportation, and live closer to where they work. If the tax were larger, more people would be driving hybrid cars, and if it were large enough, they would switch to electric cars.

When policymakers fail to consider how their policies affect incentives, they often end up with unintended consequences. For example, consider public policy regarding auto safety. Today, all cars have seat belts, but this was not true 50 years ago. In the 1960s, Ralph Nader’s book Unsafe at Any Speed generated much public concern over auto safety. Congress responded with laws requiring seat belts as standard equipment on new cars.

How does a seat belt law affect auto safety? The direct effect is obvious: When a person wears a seat belt, the probability of surviving an auto accident rises. But that’s not the end of the story because the law also affects behavior by altering incentives. The relevant behavior here is the speed and care with which drivers operate their cars. Driving slowly and carefully is costly because it uses the driver’s time and energy. When deciding how safely to drive, rational people compare, perhaps unconsciously, the marginal benefit from safer driving to the marginal cost. As result, they drive more slowly and carefully when the benefit of increased safety is high. For example, when road conditions are icy, people drive more attentively and at lower speeds than they do when road conditions are clear.

Consider how a seat belt law alters a driver’s cost–benefit calculation. Seat belts make accidents less costly because they reduce the likelihood of injury or death. In other words, seat belts reduce the benefits of slow and careful driving. People respond to seat belts as they would to an improvement in road conditions—by driving faster and less carefully. The result of a seat belt law, therefore, is a larger number of accidents. The decline in safe driving has a clear, adverse impact on pedestrians, who are more likely to find themselves in an accident but (unlike the drivers) don’t have the benefit of added protection.

At first, this discussion of incentives and seat belts might seem like idle speculation. Yet in a classic 1975 study, economist Sam Peltzman argued that auto-safety laws have had many of these effects. According to Peltzman’s evidence, these laws produce both fewer deaths per accident and more accidents. He concluded
that the net result is little change in the number of driver deaths and an increase in the number of pedestrian deaths.

Peltzman’s analysis of auto safety is an offbeat example of the general principle that people respond to incentives. When analyzing any policy, we must consider not only the direct effects but also the less obvious indirect effects that work through incentives. If the policy changes incentives, it will cause people to alter their behavior.

**QUICK QUIZ** Describe an important trade-off you recently faced. • Give an example of some action that has both a monetary and nonmonetary opportunity cost. • Describe an incentive your parents offered to you in an effort to influence your behavior.

### HOW PEOPLE INTERACT

The first four principles discussed how individuals make decisions. As we go about our lives, many of our decisions affect not only ourselves but other people as well. The next three principles concern how people interact with one another.

**PRINCIPLE 5: TRADE CAN MAKE EVERYONE BETTER OFF**

You have probably heard on the news that the Japanese are our competitors in the world economy. In some ways, this is true because American and Japanese firms produce many of the same goods. Ford and Toyota compete for the same customers in the market for automobiles. Apple and Sony compete for the same customers in the market for digital music players.

Yet it is easy to be misled when thinking about competition among countries. Trade between the United States and Japan is not like a sports contest in which one side wins and the other side loses. In fact, the opposite is true: Trade between two countries can make each country better off.

To see why, consider how trade affects your family. When a member of your family looks for a job, he or she competes against members of other families who are looking for jobs. Families also compete against one another when they go shopping because each family wants to buy the best goods at the lowest prices. In a sense, each family in the economy is competing with all other families.

Despite this competition, your family would not be better off isolating itself from all other families. If it did, your family would need to grow its own food, make its own clothes, and build its own home. Clearly, your family gains much from its ability to trade with others. Trade allows each person to specialize in the activities he or she does best, whether it is farming, sewing, or home building. By trading with others, people can buy a greater variety of goods and services at lower cost.

Countries as well as families benefit from the ability to trade with one another. Trade allows countries to specialize in what they do best and to enjoy a greater variety of goods and services. The Japanese, as well as the French and the Egyptians and the Brazilians, are as much our partners in the world economy as they are our competitors.

**PRINCIPLE 6: MARKETS ARE USUALLY A GOOD WAY TO ORGANIZE ECONOMIC ACTIVITY**

The collapse of communism in the Soviet Union and Eastern Europe in the 1980s may be the most important change in the world during the past half century.
Communist countries worked on the premise that government officials were in the best position to allocate the economy’s scarce resources. These central planners decided what goods and services were produced, how much was produced, and who produced and consumed these goods and services. The theory behind central planning was that only the government could organize economic activity in a way that promoted economic well-being for the country as a whole.

Most countries that once had centrally planned economies have abandoned the system and are instead developing market economies. In a market economy, the market allocates resources through the decentralized decisions of many firms and households as they interact in markets for goods and services.
decisions of a central planner are replaced by the decisions of millions of firms and households. Firms decide whom to hire and what to make. Households decide which firms to work for and what to buy with their incomes. These firms and households interact in the marketplace, where prices and self-interest guide their decisions.

At first glance, the success of market economies is puzzling. In a market economy, no one is looking out for the economic well-being of society as a whole. Free markets contain many buyers and sellers of numerous goods and services, and all of them are interested primarily in their own well-being. Yet despite decentralized decision making and self-interested decision makers, market economies have proven remarkably successful in organizing economic activity to promote overall economic well-being.

In his 1776 book *An Inquiry into the Nature and Causes of the Wealth of Nations*, economist Adam Smith made the most famous observation in all of economics: Households and firms interacting in markets act as if they are guided by an “invisible hand” that leads them to desirable market outcomes. One of our goals in this book is to understand how this invisible hand works its magic.

As you study economics, you will learn that prices are the instrument with which the invisible hand directs economic activity. In any market, buyers look at the price when determining how much to demand, and sellers look at the price when deciding how much to supply. As a result of the decisions that buyers and sellers make, market prices reflect both the value of a good to society and the cost to society of making the good. Smith’s great insight was that prices adjust to guide these individual buyers and sellers to reach outcomes that, in many cases, maximize the well-being of society as a whole.

Smith’s insight has an important corollary: When the government prevents prices from adjusting naturally to supply and demand, it impedes the invisible hand’s ability to coordinate the decisions of the households and firms that make up the economy. This corollary explains why taxes adversely affect the allocation of resources, for they distort prices and thus the decisions of households and firms. It also explains the great harm caused by policies that directly control prices, such as rent control. And it explains the failure of communism. In Communist countries, prices were not determined in the marketplace but were dictated by central planners. These planners lacked the necessary information about consumers’ tastes and producers’ costs, which in a market economy are reflected in prices. Central planners failed because they tried to run the economy with one hand tied behind their backs—the invisible hand of the marketplace.

**PRINCIPLE 7: GOVERNMENTS CAN SOMETIMES IMPROVE MARKET OUTCOMES**

If the invisible hand of the market is so great, why do we need government? One purpose of studying economics is to refine your view about the proper role and scope of government policy.

One reason we need government is that the invisible hand can work its magic only if the government enforces the rules and maintains the institutions that are key to a market economy. Most important, market economies need institutions to enforce property rights so individuals can own and control scarce resources. A farmer won’t grow food if he expects his crop to be stolen; a restaurant won’t serve meals unless it is assured that customers will pay before they leave; and a music company won’t produce CDs if too many potential customers avoid paying
by making illegal copies. We all rely on government-provided police and courts to enforce our rights over the things we produce—and the invisible hand counts on our ability to enforce our rights.

Yet there is another reason we need government: The invisible hand is powerful, but it is not omnipotent. There are two broad reasons for a government to intervene in the economy and change the allocation of resources that people would choose on their own: to promote efficiency or to promote equality. That is, most policies aim either to enlarge the economic pie or to change how the pie is divided.

Consider first the goal of efficiency. Although the invisible hand usually leads markets to allocate resources to maximize the size of the economic pie, this is not always the case. Economists use the term **market failure** to refer to a situation in which the market on its own fails to produce an efficient allocation of resources. As we will see, one possible cause of market failure is an **externality**, which is the impact of one person’s actions on the well-being of a bystander. The classic example of an externality is pollution. Another possible cause of market failure

---

**FYI**

**Adam Smith and the Invisible Hand**

It may be only a coincidence that Adam Smith’s great book *The Wealth of Nations* was published in 1776, the exact year American revolutionaries signed the Declaration of Independence. But the two documents share a point of view that was prevalent at the time: Individuals are usually best left to their own devices, without the heavy hand of government guiding their actions. This political philosophy provides the intellectual basis for the market economy and for free society more generally.

Why do decentralized market economies work so well? Is it because people can be counted on to treat one another with love and kindness? Not at all. Here is Adam Smith’s description of how people interact in a market economy:

*Man has almost constant occasion for the help of his brethren, and it is in vain for him to expect it from their benevolence only. He will be more likely to prevail if he can interest their self-love in his favour, and show them that it is for their own advantage to do for him what he requires of them. . . . Give me that which I want, and you shall have this which you want, is the meaning of every such offer, and it is in this manner that we obtain from one another the far greater part of those good offices which we stand in need of.*

Smith is saying that participants in the economy are motivated by self-interest and that the “invisible hand” of the marketplace guides this self-interest into promoting general economic well-being.

Many of Smith’s insights remain at the center of modern economics. Our analysis in the coming chapters will allow us to express Smith’s conclusions more precisely and to analyze more fully the strengths and weaknesses of the market’s invisible hand.
is **market power**, which refers to the ability of a single person (or small group) to unduly influence market prices. For example, if everyone in town needs water but there is only one well, the owner of the well is not subject to the rigorous competition with which the invisible hand normally keeps self-interest in check. In the presence of externalities or market power, well-designed public policy can enhance economic efficiency.

Now consider the goal of equality. Even when the invisible hand is yielding efficient outcomes, it can nonetheless leave sizable disparities in economic well-being. A market economy rewards people according to their ability to produce things that other people are willing to pay for. The world’s best basketball player earns more than the world’s best chess player simply because people are willing to pay more to watch basketball than chess. The invisible hand does not ensure that everyone has sufficient food, decent clothing, and adequate healthcare. This inequality may, depending on one’s political philosophy, call for government intervention. In practice, many public policies, such as the income tax and the welfare system, aim to achieve a more equal distribution of economic well-being.

To say that the government *can* improve on market outcomes at times does not mean that it always *will*. Public policy is made not by angels but by a political process that is far from perfect. Sometimes policies are designed simply to reward the politically powerful. Sometimes they are made by well-intentioned leaders who are not fully informed. As you study economics, you will become a better judge of when a government policy is justifiable because it promotes efficiency or equality and when it is not.

**Quick Quiz** Why is a country better off not isolating itself from all other countries?

- Why do we have markets and, according to economists, what roles should government play in them?

---

**HOW THE ECONOMY AS A WHOLE WORKS**

We started by discussing how individuals make decisions and then looked at how people interact with one another. All these decisions and interactions together make up “the economy.” The last three principles concern the workings of the economy as a whole.

**Principle 8: A Country’s Standard of Living Depends on Its Ability to Produce Goods and Services**

The differences in living standards around the world are staggering. In 2006, the average American had an income of about $44,260. In the same year, the average Mexican earned $11,410, and the average Nigerian earned $1,050. Not surprisingly, this large variation in average income is reflected in various measures of the quality of life. Citizens of high-income countries have more TV sets, more cars, better nutrition, better healthcare, and a longer life expectancy than citizens of low-income countries.

Changes in living standards over time are also large. In the United States, incomes have historically grown about 2 percent per year (after adjusting for
changes in the cost of living). At this rate, average income doubles every 35 years. Over the past century, average income has risen about eightfold.

What explains these large differences in living standards among countries and over time? The answer is surprisingly simple. Almost all variation in living standards is attributable to differences in countries’ productivity—that is, the amount of goods and services produced from each unit of labor input. In nations where workers can produce a large quantity of goods and services per unit of time, most people enjoy a high standard of living; in nations where workers are less productive, most people endure a more meager existence. Similarly, the growth rate of a nation’s productivity determines the growth rate of its average income.

The fundamental relationship between productivity and living standards is simple, but its implications are far-reaching. If productivity is the primary determinant of living standards, other explanations must be of secondary importance. For example, it might be tempting to credit labor unions or minimum-wage laws for the rise in living standards of American workers over the past century. Yet the real hero of American workers is their rising productivity. As another example, some commentators have claimed that increased competition from Japan and other countries explained the slow growth in U.S. incomes during the 1970s and 1980s. Yet the real villain was not competition from abroad but flagging productivity growth in the United States.

The relationship between productivity and living standards also has profound implications for public policy. When thinking about how any policy will affect living standards, the key question is how it will affect our ability to produce goods and services. To boost living standards, policymakers need to raise productivity by ensuring that workers are well educated, have the tools needed to produce goods and services, and have access to the best available technology.

**Principle 9: Prices Rise When the Government Prints Too Much Money**

In January 1921, a daily newspaper in Germany cost 0.30 marks. Less than two years later, in November 1922, the same newspaper cost 70,000,000 marks. All other prices in the economy rose by similar amounts. This episode is one of history’s most spectacular examples of inflation, an increase in the overall level of prices in the economy.

Although the United States has never experienced inflation even close to that in Germany in the 1920s, inflation has at times been an economic problem. During the 1970s, for instance, when the overall level of prices more than doubled, President Gerald Ford called inflation “public enemy number one.” By contrast, inflation in the first decade of the 21st century has run about 2½ percent per year; at this rate, it would take almost 30 years for prices to double. Because high inflation imposes various costs on society, keeping inflation at a low level is a goal of economic policymakers around the world.

What causes inflation? In almost all cases of large or persistent inflation, the culprit is growth in the quantity of money. When a government creates large quantities of the nation’s money, the value of the money falls. In Germany in the early 1920s, when prices were on average tripling every month, the quantity of money was also tripling every month. Although less dramatic, the economic history of the United States points to a similar conclusion: The high inflation of the 1970s was associated with rapid growth in the quantity of money, and the low
My take on training in economics is that it becomes increasingly valuable as you move up the career ladder. I can’t imagine a better major for corporate CEOs, congressmen, or American presidents. You’ve learned a systematic, disciplined way of thinking that will serve you well. By contrast, the economically challenged must be perplexed by the magic of markets and the dangers of tampering with them too much. You know better what you first learned in kindergarten: that you shouldn’t kill or cripple the goose that lays the golden eggs. . . .

Economics training will help you understand fallacies and unintended consequences. In fact, I am inclined to define economics as the study of how to anticipate unintended consequences. . . .

Little in the literature seems more relevant to contemporary economic debates about how it is that economies work better the fewer people they have in charge. Who does the planning? Who makes decisions? Who decides what to produce? For my money, Adam Smith’s invisible hand is the most important thing you’ve learned by studying economics. You understand how we can each work for our own self-interest and still produce a desirable social outcome. You know how uncoordinated activity gets coordinated by the market to enhance the wealth of nations. You understand the inflation of more recent experience was associated with slow growth in the quantity of money.

**Principle 10: Society Faces a Short-Run Trade-off between Inflation and Unemployment**

Although a higher level of prices is, in the long run, the primary effect of increasing the quantity of money, the short-run story is more complex and controversial. Most economists describe the short-run effects of monetary injections as follows:

- Increasing the amount of money in the economy stimulates the overall level of spending and thus the demand for goods and services.
- Higher demand may over time cause firms to raise their prices, but in the meantime, it also encourages them to hire more workers and produce a larger quantity of goods and services.
- More hiring means lower unemployment.

This line of reasoning leads to one final economy-wide trade-off: a short-run trade-off between inflation and unemployment. Although some economists still question these ideas, most accept that society faces a short-run trade-off between inflation and unemployment. This simply means that, over a period of a year or two, many economic policies push inflation and unemployment in opposite directions. Policymakers face this trade-off regardless of whether inflation and unemployment both start out at high levels (as they were in the early 1980s), at low levels (as they were in the late 1990s),
than what usually is called the broken window fallacy. Whenever a government program is justified not on its merits but by the jobs it will create, remember the broken window: Some teenagers, being the little beasts that they are, toss a brick through a bakery window. A crowd gathers and laments, “What a shame.” But before you know it, someone suggests a silver lining to the situation: Now the baker will have to spend money to have the window repaired. This will add to the income of the repairman, who will spend his additional income, which will add to another seller’s income, and so on. You know the drill. The chain of spending will multiply and generate higher income and employment. If the broken window is large enough, it might produce an economic boom!...

Most voters fall for the broken window fallacy, but not economics majors. They will say, “Hey, wait a minute!” If the baker hadn’t spent his money on window repair, he would have spent it on the new suit he was saving to buy. Then the tailor would have the new income to spend, and so on. The broken window didn’t create net new spending; it just diverted spending from somewhere else. The broken window does not create new activity, just different activity. People see the activity that takes place. They don’t see the activity that would have taken place.

The broken window fallacy is perpetuated in many forms. Whenever job creation or retention is the primary objective I call it the job-counting fallacy. Economics majors understand the non-intuitive reality that real progress comes from job destruction. It once took 90 percent of our population to grow our food. Now it takes 3 percent. Pardon me, Willie, but are we worse off because of the job losses in agriculture? The would-have-been farmers are now college professors and computer gurus...

So instead of counting jobs, we should make every job count. We will occasionally hit a soft spot when we have a mismatch of supply and demand in the labor market. But that is temporary. Don’t become a Luddite and destroy the machinery, or become a protectionist and try to grow bananas in New York City.


or someplace in between. This short-run trade-off plays a key role in the analysis of the business cycle—the irregular and largely unpredictable fluctuations in economic activity, as measured by the production of goods and services or the number of people employed.

Policymakers can exploit the short-run trade-off between inflation and unemployment using various policy instruments. By changing the amount that the government spends, the amount it taxes, and the amount of money it prints, policymakers can influence the overall demand for goods and services. Changes in demand in turn influence the combination of inflation and unemployment that the economy experiences in the short-run. Because these instruments of economic policy are potentially so powerful, how policymakers should use these instruments to control the economy, if at all, is a subject of continuing debate.

Quick Quiz List and briefly explain the three principles that describe how the economy as a whole works.

Conclusion You now have a taste of what economics is all about. In the coming chapters, we develop many specific insights about people, markets, and economies. Mastering these insights will take some effort, but it is not an overwhelming task. The field of economics is based on a few big ideas that can be applied in many different situations.
How to Read This Book

Economics is fun, but it can also be hard to learn. My aim in writing this text is to make it as enjoyable and easy as possible. But you, the student, also have a role to play. Experience shows that if you are actively involved as you study this book, you will enjoy a better outcome both on your exams and in the years that follow. Here are a few tips about how best to read this book.

1. **Read before class.** Students do better when they read the relevant textbook chapter before attending a lecture. You will understand the lecture better, and your questions will be better focused on where you need extra help.

2. **Summarize, don’t highlight.** Running a yellow marker over the text is too passive an activity to keep your mind engaged. Instead, when you come to the end of a section, take a minute and summarize what you just learned in your own words, writing your summary in the wide margins we’ve provided. When you’ve finished the chapter, compare your summaries with the one at the end of the chapter. Did you pick up the main points?

3. **Test yourself.** Throughout the book, Quick Quizzes offer instant feedback to find out if you’ve learned what you are supposed to. Take the opportunity to write down your answer and then check it against the answers provided at this book’s website. The quizzes are meant to test your basic comprehension. If your answer is incorrect, you probably need to review the section.

4. **Practice, practice, practice.** At the end of each chapter, Questions for Review test your understanding, and Problems and Applications ask you to apply and extend the material. Perhaps your instructor will assign some of these exercises as homework. If so, do them. If not, do them anyway. The more you use your new knowledge, the more solid it becomes.

5. **Go online.** The publisher of this book maintains an extensive website to help you in your study of economics. It includes additional examples, applications, and problems, as well as quizzes so you can test yourself. Check it out. The website is http://academic.cengage.com/economics/mankiw.

6. **Study in groups.** After you’ve read the book and worked problems on your own, get together with classmates to discuss the material. You will learn from each other—an example of the gains from trade.

7. **Teach someone.** As all teachers know, there is no better way to learn something than to teach it to someone else. Take the opportunity to teach new economic concepts to a study partner, a friend, a parent, or even a pet.

8. **Don’t skip the real-world examples.** In the midst of all the numbers, graphs, and strange new words, it is easy to lose sight of what economics is all about. The Case Studies and In the News boxes sprinkled throughout this book should help remind you. They show how the theory is tied to events happening in all our lives.

9. **Apply economic thinking to your daily life.** Once you’ve read about how others apply economics to the real world, try it yourself! You can use economic analysis to better understand your own decisions, the economy around you, and the events you read about in the newspaper. The world may never look the same again.

Throughout this book, we will refer back to the Ten Principles of Economics highlighted in this chapter and summarized in Table 1. Keep these building blocks in mind: Even the most sophisticated economic analysis is founded on the ten principles introduced here.
How People Make Decisions
1: People Face Trade-offs
2: The Cost of Something Is What You Give Up to Get It
3: Rational People Think at the Margin
4: People Respond to Incentives

How People Interact
5: Trade Can Make Everyone Better Off
6: Markets Are Usually a Good Way to Organize Economic Activity
7: Governments Can Sometimes Improve Market Outcomes

How the Economy as a Whole Works
8: A Country’s Standard of Living Depends on Its Ability to Produce Goods and Services
9: Prices Rise When the Government Prints Too Much Money
10: Society Faces a Short-Run Trade-off between Inflation and Unemployment

SUMMARY
- The fundamental lessons about individual decision making are that people face trade-offs among alternative goals, that the cost of any action is measured in terms of forgone opportunities, that rational people make decisions by comparing marginal costs and marginal benefits, and that people change their behavior in response to the incentives they face.
- The fundamental lessons about interactions among people are that trade and interdependence can be mutually beneficial, that markets are usually a good way of coordinating economic activity among people, and that the government can potentially improve market outcomes by remedying a market failure or by promoting greater economic equality.
- The fundamental lessons about the economy as a whole are that productivity is the ultimate source of living standards, that growth in the quantity of money is the ultimate source of inflation, and that society faces a short-run trade-off between inflation and unemployment.

KEY CONCEPTS
scarcity, p. 3
economics, p. 4
efficiency, p. 5
equality, p. 5
opportunity cost, p. 5
rational people, p. 6
marginal changes, p. 6
incentive, p. 7
market economy, p. 9
property rights, p. 10
market failure, p. 11
externality, p. 11
market power, p. 12
productivity, p. 13
inflation, p. 13
business cycle, p. 15
QUESTIONS FOR REVIEW

1. Give three examples of important trade-offs that you face in your life.
2. What is the opportunity cost of seeing a movie?
3. Water is necessary for life. Is the marginal benefit of a glass of water large or small?
4. Why should policymakers think about incentives?
5. Why isn’t trade among countries like a game with some winners and some losers?
6. What does the “invisible hand” of the marketplace do?
7. Explain the two main causes of market failure and give an example of each.
8. Why is productivity important?
9. What is inflation and what causes it?
10. How are inflation and unemployment related in the short run?

PROBLEMS AND APPLICATIONS

1. Describe some of the trade-offs faced by each of the following:
   a. a family deciding whether to buy a new car
   b. a member of Congress deciding how much to spend on national parks
   c. a company president deciding whether to open a new factory
   d. a professor deciding how much to prepare for class
   e. a recent college graduate deciding whether to go to graduate school
2. You are trying to decide whether to take a vacation. Most of the costs of the vacation (airfare, hotel, and forgone wages) are measured in dollars, but the benefits of the vacation are psychological. How can you compare the benefits to the costs?
3. You were planning to spend Saturday working at your part-time job, but a friend asks you to go skiing. What is the true cost of going skiing? Now suppose you had been planning to spend the day studying at the library. What is the cost of going skiing in this case? Explain.
4. You win $100 in a basketball pool. You have a choice between spending the money now or putting it away for a year in a bank account that pays 5 percent interest. What is the opportunity cost of spending the $100 now?
5. The company that you manage has invested $5 million in developing a new product, but the development is not quite finished. At a recent meeting, your salespeople report that the introduction of competing products has reduced the expected sales of your new product to $3 million. If it would cost $1 million to finish development and make the product, should you go ahead and do so? What is the most that you should pay to complete development?
6. Three managers of the Magic Potion Company are discussing a possible increase in production. Each suggests a way to make this decision.
   - **Harry:** We should examine whether our company’s productivity—gallons of potion per worker—would rise or fall.
   - **Ron:** We should examine whether our average cost—cost per worker—would rise or fall.
   - **Hermione:** We should examine whether the extra revenue from selling the additional potion would be greater or smaller than the extra costs.
   Who do you think is right? Why?
7. The Social Security system provides income for people over age 65. If a recipient of Social Security decides to work and earn some income, the amount he or she receives in Social Security benefits is typically reduced.
   a. How does the provision of Social Security affect people’s incentive to save while working?
   b. How does the reduction in benefits associated with higher earnings affect people’s incentive to work past age 65?
8. A recent bill reforming the government’s anti-poverty programs limited many welfare recipients to only two years of benefits.
   a. How does this change affect the incentives for working?
   b. How might this change represent a trade-off between equality and efficiency?
9. Your roommate is a better cook than you are, but you can clean more quickly than your roommate can. If your roommate did all the cooking and you did all the cleaning, would your chores take you more or less time than if you divided each task evenly? Give a similar example of how specialization and trade can make two countries both better off.
10. Suppose the United States adopted central planning for its economy, and you became the chief planner. Among the millions of decisions that you need to make for next year are how many compact discs to produce, what artists to record, and which consumers should receive the discs. To make these decisions intelligently, what information would you need about the compact disc industry? What information would you need about each of the people in the United States? How well do you think you could do your job?
11. Explain whether each of the following government activities is motivated by a concern about equality or a concern about efficiency. In the case of efficiency, discuss the type of market failure involved.
   a. regulating cable TV prices
   b. providing some poor people with vouchers that can be used to buy food
   c. prohibiting smoking in public places
   d. breaking up Standard Oil (which once owned 90 percent of all oil refineries) into several smaller companies
   e. imposing higher personal income tax rates on people with higher incomes
   f. instituting laws against driving while intoxicated
12. Discuss each of the following statements from the standpoints of equality and efficiency.
   a. “Everyone in society should be guaranteed the best healthcare possible.”
   b. “When workers are laid off, they should be able to collect unemployment benefits until they find a new job.”
13. In what ways is your standard of living different from that of your parents or grandparents when they were your age? Why have these changes occurred?
14. Suppose Americans decide to save more of their incomes. If banks lend this extra savings to businesses, which use the funds to build new factories, how might this lead to faster growth in productivity? Who do you suppose benefits from the higher productivity? Is society getting a free lunch?
15. During the Revolutionary War, the American colonies could not raise enough tax revenue to fully fund the war effort; to make up this difference, the colonies decided to print more money. Printing money to cover expenditures is sometimes referred to as an “inflation tax.” Who do you think is being “taxed” when more money is printed? Why?
16. Imagine that you are a policymaker trying to decide whether to reduce the rate of inflation. To make an intelligent decision, what would you need to know about inflation, unemployment, and the trade-off between them?
This page intentionally left blank
Thinking Like an Economist

Every field of study has its own language and its own way of thinking. Mathematicians talk about axioms, integrals, and vector spaces. Psychologists talk about ego, id, and cognitive dissonance. Lawyers talk about venue, torts, and promissory estoppel.

Economics is no different. Supply, demand, elasticity, comparative advantage, consumer surplus, deadweight loss—these terms are part of the economist’s language. In the coming chapters, you will encounter many new terms and some familiar words that economists use in specialized ways. At first, this new language may seem needlessly arcane. But as you will see, its value lies in its ability to provide you with a new and useful way of thinking about the world in which you live.

The purpose of this book is to help you learn the economist’s way of thinking. Just as you cannot become a mathematician, psychologist, or lawyer overnight, learning to think like an economist will take some time. Yet with a combination of theory, case studies, and examples of economics in the news, this book will give you ample opportunity to develop and practice this skill.

Before delving into the substance and details of economics, it is helpful to have an overview of how economists approach the world. This chapter discusses the field’s methodology. What is distinctive about how economists confront a question? What does it mean to think like an economist?
Economists try to address their subject with a scientist’s objectivity. They approach the study of the economy in much the same way a physicist approaches the study of matter and a biologist approaches the study of life: They devise theories, collect data, and then analyze these data in an attempt to verify or refute their theories.

To beginners, it can seem odd to claim that economics is a science. After all, economists do not work with test tubes or telescopes. The essence of science, however, is the scientific method—the dispassionate development and testing of theories about how the world works. This method of inquiry is as applicable to studying a nation’s economy as it is to studying the earth’s gravity or a species’ evolution. As Albert Einstein once put it, “The whole of science is nothing more than the refinement of everyday thinking.”

Although Einstein’s comment is as true for social sciences such as economics as it is for natural sciences such as physics, most people are not accustomed to looking at society through the eyes of a scientist. Let’s discuss some of the ways in which economists apply the logic of science to examine how an economy works.

**THE SCIENTIFIC METHOD: OBSERVATION, THEORY, AND MORE OBSERVATION**

Isaac Newton, the famous 17th-century scientist and mathematician, allegedly became intrigued one day when he saw an apple fall from a tree. This observation motivated Newton to develop a theory of gravity that applies not only to an apple falling to the earth but to any two objects in the universe. Subsequent testing of Newton’s theory has shown that it works well in many circumstances (although, as Einstein would later emphasize, not in all circumstances). Because Newton’s theory has been so successful at explaining observation, it is still taught in undergraduate physics courses around the world.

This interplay between theory and observation also occurs in the field of economics. An economist might live in a country experiencing rapidly increasing prices and be moved by this observation to develop a theory of inflation. The theory might assert that high inflation arises when the government prints too much money. To test this theory, the economist could collect and analyze data on prices and money from many different countries. If growth in the quantity of money were not at all related to the rate at which prices are rising, the economist would start to doubt the validity of this theory of inflation. If money growth and inflation were strongly correlated in international data, as in fact they are, the economist would become more confident in the theory.

Although economists use theory and observation like other scientists, they face an obstacle that makes their task especially challenging: In economics, conducting experiments is often difficult and sometimes impossible. Physicists studying gravity can drop many objects in their laboratories to generate data to test their theories. By contrast, economists studying inflation are not allowed to manipulate a nation’s monetary policy simply to generate useful data. Economists, like astronomers and evolutionary biologists, usually have to make do with whatever data the world happens to give them.

To find a substitute for laboratory experiments, economists pay close attention to the natural experiments offered by history. When a war in the Middle East interrupts the flow of crude oil, for instance, oil prices skyrocket around the world.
For consumers of oil and oil products, such an event depresses living standards. For economic policymakers, it poses a difficult choice about how best to respond. But for economic scientists, the event provides an opportunity to study the effects of a key natural resource on the world’s economies. Throughout this book, therefore, we consider many historical episodes. These episodes are valuable to study because they give us insight into the economy of the past and, more important, because they allow us to illustrate and evaluate economic theories of the present.

**The Role of Assumptions**

If you ask a physicist how long it would take a marble to fall from the top of a ten-story building, she will likely answer the question by assuming that the marble falls in a vacuum. Of course, this assumption is false. In fact, the building is surrounded by air, which exerts friction on the falling marble and slows it down. Yet the physicist will point out that friction on the marble is so small that its effect is negligible. Assuming the marble falls in a vacuum simplifies the problem without substantially affecting the answer.

Economists make assumptions for the same reason: Assumptions can simplify the complex world and make it easier to understand. To study the effects of international trade, for example, we might assume that the world consists of only two countries and that each country produces only two goods. In reality, there are numerous countries, each of which produces thousands of different types of goods. But by assuming two countries and two goods, we can focus our thinking on the essence of the problem. Once we understand international trade in this simplified imaginary world, we are in a better position to understand international trade in the more complex world in which we live.

The art in scientific thinking—whether in physics, biology, or economics—is deciding which assumptions to make. Suppose, for instance, that we were dropping a beachball rather than a marble from the top of the building. Our physicist would realize that the assumption of no friction is less accurate in this case: Friction exerts a greater force on a beachball than on a marble because a beachball is much larger. The assumption that gravity works in a vacuum is reasonable for studying a falling marble but not for studying a falling beachball.

Similarly, economists use different assumptions to answer different questions. Suppose that we want to study what happens to the economy when the government changes the number of dollars in circulation. An important piece of this analysis, it turns out, is how prices respond. Many prices in the economy change infrequently; the newstand prices of magazines, for instance, change only every few years. Knowing this fact may lead us to make different assumptions when studying the effects of the policy change over different time horizons. For studying the short-run effects of the policy, we may assume that prices do not change much. We may even make the extreme and artificial assumption that all prices are completely fixed. For studying the long-run effects of the policy, however, we may assume that all prices are completely flexible. Just as a physicist uses different assumptions when studying falling marbles and falling beachballs, economists use different assumptions when studying the short-run and long-run effects of a change in the quantity of money.

**Economic Models**

High school biology teachers teach basic anatomy with plastic replicas of the human body. These models have all the major organs: the heart, the liver, the
kidneys, and so on. The models allow teachers to show their students very simply how the important parts of the body fit together. Because these plastic models are stylized and omit many details, no one would mistake one of them for a real person. Despite this lack of realism—indeed, because of this lack of realism—studying these models is useful for learning how the human body works.

Economists also use models to learn about the world, but instead of being made of plastic, they are most often composed of diagrams and equations. Like a biology teacher’s plastic model, economic models omit many details to allow us to see what is truly important. Just as the biology teacher’s model does not include all the body’s muscles and capillaries, an economist’s model does not include every feature of the economy.

As we use models to examine various economic issues throughout this book, you will see that all the models are built with assumptions. Just as a physicist begins the analysis of a falling marble by assuming away the existence of friction, economists assume away many of the details of the economy that are irrelevant for studying the question at hand. All models—in physics, biology, and economics—simplify reality to improve our understanding of it.

**OUR FIRST MODEL: THE CIRCULAR-FLOW DIAGRAM**

The economy consists of millions of people engaged in many activities—buying, selling, working, hiring, manufacturing, and so on. To understand how the economy works, we must find some way to simplify our thinking about all these activities. In other words, we need a model that explains, in general terms, how the economy is organized and how participants in the economy interact with one another.

Figure 1 presents a visual model of the economy called a **circular-flow diagram**. In this model, the economy is simplified to include only two types of decision makers—firms and households. Firms produce goods and services using inputs, such as labor, land, and capital (buildings and machines). These inputs are called the **factors of production**. Households own the factors of production and consume all the goods and services that the firms produce.

Households and firms interact in two types of markets. In the **markets for goods and services**, households are buyers, and firms are sellers. In particular, households buy the output of goods and services that firms produce. In the **markets for the factors of production**, households are sellers, and firms are buyers. In these markets, households provide the inputs that firms use to produce goods and services. The circular-flow diagram offers a simple way of organizing the economic transactions that occur between households and firms in the economy.

The two loops of the circular-flow diagram are distinct but related. The inner loop represents the flows of inputs and outputs. The households sell the use of their labor, land, and capital to the firms in the markets for the factors of production. The firms then use these factors to produce goods and services, which in turn are sold to households in the markets for goods and services. The outer loop of the diagram represents the corresponding flow of dollars. The households spend money to buy goods and services from the firms. The firms use some of the revenue from these sales to pay for the factors of production, such as the wages of their workers. What’s left is the profit of the firm owners, who themselves are members of households.

Let’s take a tour of the circular flow by following a dollar bill as it makes its way from person to person through the economy. Imagine that the dollar begins at a household, say, in your wallet. If you want to buy a cup of coffee, you take the dollar to one of the economy’s markets for goods and services, such as your local
Starbucks coffee shop. There you spend it on your favorite drink. When the dollar moves into the Starbucks cash register, it becomes revenue for the firm. The dollar doesn’t stay at Starbucks for long, however, because the firm uses it to buy inputs in the markets for the factors of production. Starbucks might use the dollar to pay rent to its landlord for the space it occupies or to pay the wages of its workers. In either case, the dollar enters the income of some household and, once again, is back in someone’s wallet. At that point, the story of the economy’s circular flow starts once again.

The circular-flow diagram in Figure 1 is one simple model of the economy. It dispenses with details that, for some purposes, are significant. A more complex and realistic circular-flow model would include, for instance, the roles of government and international trade. (Some of that dollar you gave to Starbucks might be used to pay taxes and or to buy coffee beans from a farmer in Brazil.) Yet these details are not crucial for a basic understanding of how the economy is organized. Because of its simplicity, this circular-flow diagram is useful to keep in mind when thinking about how the pieces of the economy fit together.

**OUR SECOND MODEL: THE PRODUCTION POSSIBILITIES FRONTIER**

Most economic models, unlike the circular-flow diagram, are built using the tools of mathematics. Here we use one of the simplest such models, called the production possibilities frontier, to illustrate some basic economic ideas.
Although real economies produce thousands of goods and services, let’s assume an economy that produces only two goods—cars and computers. Together, the car industry and the computer industry use all of the economy’s factors of production. The production possibilities frontier is a graph that shows the various combinations of output—in this case, cars and computers—that the economy can possibly produce given the available factors of production and the available production technology that firms use to turn these factors into output.

Figure 2 shows this economy’s production possibilities frontier. If the economy uses all its resources in the car industry, it produces 1,000 cars and no computers. If it uses all its resources in the computer industry, it produces 3,000 computers and no cars. The two endpoints of the production possibilities frontier represent these extreme possibilities.

More likely, the economy divides its resources between the two industries, and this yields other points on the production possibilities frontier. For example, it can produce 600 cars and 2,200 computers, shown in the figure by point A. Or, by moving some of the factors of production to the car industry from the computer industry, the economy can produce 700 cars and 2,000 computers, represented by point B.

Because resources are scarce, not every conceivable outcome is feasible. For example, no matter how resources are allocated between the two industries, the economy cannot produce the amount of cars and computers represented by point C. Given the technology available for manufacturing cars and computers, the economy does not have enough of the factors of production to support that level of output. With the resources it has, the economy can produce at any point on or inside the production possibilities frontier, but it cannot produce at points outside the frontier.

An outcome is said to be efficient if the economy is getting all it can from the scarce resources it has available. Points on (rather than inside) the production possibilities frontier represent efficient levels of production. When the economy is producing at such a point, say point A, there is no way to produce more of one
good without producing less of the other. Point D represents an *inefficient* outcome. For some reason, perhaps widespread unemployment, the economy is producing less than it could from the resources it has available: It is producing only 300 cars and 1,000 computers. If the source of the inefficiency is eliminated, the economy can increase its production of both goods. For example, if the economy moves from point D to point A, its production of cars increases from 300 to 600, and its production of computers increases from 1,000 to 2,200.

One of the *Ten Principles of Economics* discussed in Chapter 1 is that people face trade-offs. The production possibilities frontier shows one trade-off that society faces. Once we have reached the efficient points on the frontier, the only way of getting more of one good is to get less of the other. When the economy moves from point A to point B, for instance, society produces 100 more cars but at the expense of producing 200 fewer computers.

This trade-off helps us understand another of the *Ten Principles of Economics*: The cost of something is what you give up to get it. This is called the *opportunity cost*. The production possibilities frontier shows the opportunity cost of one good as measured in terms of the other good. When society moves from point A to point B, it gives up 200 computers to get 100 additional cars. That is, at point A, the opportunity cost of 100 cars is 200 computers. Put another way, the opportunity cost of each car is two computers. Notice that the opportunity cost of a car equals the slope of the production possibilities frontier. (If you don’t recall what slope is, you can refresh your memory with the graphing appendix to this chapter.)

The opportunity cost of a car in terms of the number of computers is not constant in this economy but depends on how many cars and computers the economy is producing. This is reflected in the shape of the production possibilities frontier. Because the production possibilities frontier in Figure 2 is bowed outward, the opportunity cost of a car is highest when the economy is producing many cars and fewer computers, such as at point E, where the frontier is steep. When the economy is producing few cars and many computers, such as at point F, the frontier is flatter, and the opportunity cost of a car is lower.

Economists believe that production possibilities frontiers often have this bowed shape. When the economy is using most of its resources to make computers, such as at point F, the resources best suited to car production, such as skilled autoworkers, are being used in the computer industry. Because these workers probably aren’t very good at making computers, the economy won’t have to lose much computer production to increase car production by one unit. The opportunity cost of a car in terms of computers is small, and the frontier is relatively flat. By contrast, when the economy is using most of its resources to make cars, such as at point E, the resources best suited to making cars are already in the car industry. Producing an additional car means moving some of the best computer technicians out of the computer industry and making them autoworkers. As a result, producing an additional car will mean a substantial loss of computer output. The opportunity cost of a car is high, and the frontier is steep.

The production possibilities frontier shows the trade-off between the outputs of different goods at a given time, but the trade-off can change over time. For example, suppose a technological advance in the computer industry raises the number of computers that a worker can produce per week. This advance expands society’s set of opportunities. For any given number of cars, the economy can make more computers. If the economy does not produce any computers, it can still produce 1,000 cars, so one endpoint of the frontier stays the same. But the rest of the production possibilities frontier shifts outward, as in Figure 3.
This figure illustrates economic growth. Society can move production from a point on the old frontier to a point on the new frontier. Which point it chooses depends on its preferences for the two goods. In this example, society moves from point A to point G, enjoying more computers (2,300 instead of 2,200) and more cars (650 instead of 600).

The production possibilities frontier simplifies a complex economy to highlight some basic but powerful ideas: scarcity, efficiency, trade-offs, opportunity cost, and economic growth. As you study economics, these ideas will recur in various forms. The production possibilities frontier offers one simple way of thinking about them.

**Microeconomics and Macroeconomics**

Many subjects are studied on various levels. Consider biology, for example. Molecular biologists study the chemical compounds that make up living things. Cellular biologists study cells, which are made up of many chemical compounds and, at the same time, are themselves the building blocks of living organisms. Evolutionary biologists study the many varieties of animals and plants and how species change gradually over the centuries.

Economics is also studied on various levels. We can study the decisions of individual households and firms. Or we can study the interaction of households and firms in markets for specific goods and services. Or we can study the operation of the economy as a whole, which is the sum of the activities of all these decision makers in all these markets.

The field of economics is traditionally divided into two broad subfields. **Microeconomics** is the study of how households and firms make decisions and how they interact in specific markets. **Macroeconomics** is the study of economy-wide phenomena. A microeconomist might study the effects of rent control on
housing in New York City, the impact of foreign competition on the U.S. auto
industry, or the effects of compulsory school attendance on workers’ earnings. A
macroeconomist might study the effects of borrowing by the federal government,
the changes over time in the economy’s rate of unemployment, or alternative poli-
cies to promote growth in national living standards.

Microeconomics and macroeconomics are closely intertwined. Because changes
in the overall economy arise from the decisions of millions of individuals, it is
impossible to understand macroeconomic developments without considering the
associated microeconomic decisions. For example, a macroeconomist might study
the effect of a federal income tax cut on the overall production of goods and ser-
vice. But to analyze this issue, he or she must consider how the tax cut affects the
decisions of households about how much to spend on goods and services.

Who Studies Economics?

As a college student, you
might be asking yourself: How many economics classes should I
take? How useful will this stuff be to me later in life? Economics can
seem abstract at first, but the field is fundamentally very practical,
and the study of economics is useful in many different career paths.
Here is a small sampling of some well-known people who majored
in economics when they were in college.

Meg Whitman  President and Chief Executive Officer, eBay
Ronald Reagan  Former President of the United States
William F. Buckley Jr.  Journalist
Danny Glover  Actor
Barbara Boxer  U.S. Senator
John Elway  NFL Quarterback
Kofi Annan  Former Secretary General, United Nations
Ted Turner  Founder of CNN and Owner of Atlanta Braves
Lionel Richie  Singer
Diane von Furstenberg  Fashion Designer
Michael Kinsley  Journalist
Ben Stein  Political Speechwriter, Actor, and Game Show Host
Cate Blanchett  Actor
Anthony Zinni  General (ret.), U.S. Marine Corps
Tiger Woods  Golfer
Steve Ballmer  Chief Executive Officer, Microsoft

Arnold Schwarzenegger  Governor of California
Sandra Day-O’Connor  Former Supreme Court Justice
Scott Adams  Cartoonist
Mick Jagger  Singer for The Rolling Stones

Having studied at the London School of Economics may not help
Mick Jagger hit the high notes, but it has probably given him some
insight about how to invest the substantial sums he has earned dur-
ing his rock-'n'-roll career.

When asked in 2005 why The Rolling Stones were going on tour again, former econom-
ics major Mick Jagger replied, “Supply and demand.” Keith Richards added, “If the
demand’s there, we’ll supply.”
Despite the inherent link between microeconomics and macroeconomics, the two fields are distinct. Because they address different questions, each field has its own set of models, which are often taught in separate courses.

**Quick Quiz** In what sense is economics like a science? • Draw a production possibilities frontier for a society that produces food and clothing. Show an efficient point, an inefficient point, and an infeasible point. Show the effects of a drought. • Define *microeconomics* and *macroeconomics*.

---

**The Economist as Policy Adviser**

Often, economists are asked to explain the causes of economic events. Why, for example, is unemployment higher for teenagers than for older workers? Sometimes, economists are asked to recommend policies to improve economic outcomes. What, for instance, should the government do to improve the economic well-being of teenagers? When economists are trying to explain the world, they are scientists. When they are trying to help improve it, they are policy advisers.

**Positive versus Normative Analysis**

To help clarify the two roles that economists play, let’s examine the use of language. Because scientists and policy advisers have different goals, they use language in different ways.

For example, suppose that two people are discussing minimum-wage laws. Here are two statements you might hear:

**Polly:** Minimum-wage laws cause unemployment.

**Norm:** The government should raise the minimum wage.

Ignoring for now whether you agree with these statements, notice that Polly and Norm differ in what they are trying to do. Polly is speaking like a scientist: She is making a claim about how the world works. Norm is speaking like a policy adviser: He is making a claim about how he would like to change the world.

In general, statements about the world are of two types. One type, such as Polly’s, is positive. **Positive statements** are descriptive. They make a claim about how the world *is*. A second type of statement, such as Norm’s, is normative. **Normative statements** are prescriptive. They make a claim about how the world *ought to be*.

A key difference between positive and normative statements is how we judge their validity. We can, in principle, confirm or refute positive statements by examining evidence. An economist might evaluate Polly’s statement by analyzing data on changes in minimum wages and changes in unemployment over time. By contrast, evaluating normative statements involves values as well as facts. Norm’s statement cannot be judged using data alone. Deciding what is good or bad policy is not just a matter of science. It also involves our views on ethics, religion, and political philosophy.

Positive and normative statements are fundamentally different, but they are often intertwined in a person’s set of beliefs. In particular, positive views about how the world works affect normative views about what policies are desirable. Polly’s claim that the minimum wage causes unemployment, if true, might lead
her to reject Norm’s conclusion that the government should raise the minimum wage. Yet normative conclusions cannot come from positive analysis alone; they involve value judgments as well.

As you study economics, keep in mind the distinction between positive and normative statements because it will help you stay focused on the task at hand. Much of economics is positive: It just tries to explain how the economy works. Yet those who use economics often have normative goals: They want to learn how to improve the economy. When you hear economists making normative statements, you know they are speaking not as scientists but as policy advisers.

**ECONOMISTS IN WASHINGTON**

President Harry Truman once said that he wanted to find a one-armed economist. When he asked his economists for advice, they always answered, “On the one hand, . . . On the other hand, . . .”

Truman was right in realizing that economists’ advice is not always straightforward. This tendency is rooted in one of the Ten Principles of Economics: People face trade-offs. Economists are aware that trade-offs are involved in most policy decisions. A policy might increase efficiency at the cost of equality. It might help future generations but hurt current generations. An economist who says that all policy decisions are easy is an economist not to be trusted.

Truman was also not alone among presidents in relying on the advice of economists. Since 1946, the president of the United States has received guidance from the Council of Economic Advisers, which consists of three members and a staff of several dozen economists. The council, whose offices are just a few steps from the White House, has no duty other than to advise the president and to write the annual Economic Report of the President, which discusses recent developments in the economy and presents the council’s analysis of current policy issues.

The president also receives input from economists in many administrative departments. Economists at the Department of the Treasury help design tax policy. Economists at the Department of Labor analyze data on workers and those looking for work to help formulate labor-market policies. Economists at the Department of Justice help enforce the nation’s antitrust laws.

Economists are also found outside the administrative branch of government. To obtain independent evaluations of policy proposals, Congress relies on the advice of the Congressional Budget Office, which is staffed by economists. The Federal Reserve, the institution that sets the nation’s monetary policy, employs hundreds of economists to analyze economic developments in the United States and throughout the world.

The influence of economists on policy goes beyond their role as advisers: Their research and writings often affect policy indirectly. Economist John Maynard Keynes offered this observation:

> The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed, the world is ruled by little else. Practical men, who believe themselves to be quite exempt from intellectual influences, are usually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back.

Although these words were written in 1935, they remain true. Indeed, the “academic scribbler” now influencing public policy is often Keynes himself.
WHY ECONOMISTS’ ADVICE IS NOT ALWAYS FOLLOWED

Any economist who advises presidents or other elected leaders knows that his or her recommendations are not always heeded. Frustrating as this can be, it is easy to understand. The process by which economic policy is made differs in many ways from the idealized policy process assumed in economics textbooks.

Throughout this text, whenever we discuss economic policy, we often focus on one question: What is the best policy for the government to pursue? We act as if policy were set by a benevolent king. Once the king figures out the right policy, he has no trouble putting his ideas into action.

In the real world, figuring out the right policy is only part of a leader’s job, sometimes the easiest part. After a president hears from his economic advisers about what policy is best from their perspective, he turns to other advisers for related input. His communications advisers will tell him how best to explain the proposed policy to the public, and they will try to anticipate any misunderstand-

GO FOR IT ON FOURTH DOWN, COACH? MAYBE YOU SHOULD ASK AN EGGHEAD.

By Shankar Vedantam

With just over five minutes to play in yesterday’s game against the New York Jets, the Washington Redskins found themselves on their own 23-yard line facing a fourth and one. The team, which was ahead by just three points, elected to do what teams normally do in such situations: They played it safe and punted rather than try to keep the drive alive.

The Jets promptly came back to kick a field goal, tying the game and sending it into overtime. While this particular story had a happy ending for Washington, which won, 23–20, it illustrated the value of an analysis by David Romer, an economist at the University of California, who has concluded that football teams are far too conservative in play calling in fourth-down situations.

You don’t have to be particularly interested in sports to find Romer’s conclusion intriguing: His hunch about human behavior in general was that although people say they have a certain goal and are willing to do everything they can to achieve it, their actual behavior regularly departs from the optimal path to reach that goal.

In his analysis of football teams, Romer specifically looked at a single question—whether teams should punt or kick the football on fourth down, or take a chance and run or throw the ball. Romer’s calculations don’t necessarily tell teams what to do in specific situations such as yesterday’s game. But on average, teams that take the risk seem to win more often than lose.

Data from a large number of NFL games show that coaches rarely follow what Romer’s calculations predict would give them the best chance of victory. While fans often suggest more aggressive play calling, even fans usually don’t go as far as the economist does—his calculations show that teams should regularly be going for it on fourth down, even if it is early in the game.
even if the score is tied, and even if the ball is on their own side of the field.

Romer’s calculations have been backed up by independent analyses. Coaches have not raised a serious challenge to Romer’s analysis, but they have simply ignored his finding.

New England Patriots coach Bill Belichick is among those who has said he agrees with Romer, and Belichick happens to be one of the more successful coaches in the league. Two Sundays ago, as the Patriots were piling up an astronomical score against Washington, Belichick took a chance on a fourth-down play and got his team seven points instead of the three he might have gotten had the team tried a field goal.

When asked by reporters why he took the chance, Belichick’s response was the response of someone who really means what he says about maximizing points: “What do you want us to do, kick a field goal?”

Owners and fans have been receptive to Romer’s ideas. However, in informal conversations Romer has had with the coaching staffs of various teams, the economist said he has been told to mind his own business in the ivory tower.

Indeed, since Romer wrote his paper a couple of years ago, NFL coaches seem to have gotten even more conservative in their play calling, which the economist attributes to their unwillingness to follow the advice of an academic, however useful it might be.

“It used to be that going for it on fourth down was the macho thing to do,” Romer said. But after his findings were widely publicized in sports circles, he said: “Now going for it on fourth down is the egghead thing to do. Would you rather be macho or an egghead?”

The interesting question raised by Romer’s research applies to a range of settings that have nothing to do with sports. Why do coaches persist in doing something that is less than optimal, when they say their only goal is to win? One theory that Romer has heard is that coaches—like generals, stock fund directors and managers in general—actually have different goals than the people they lead and the people they must answer to. Everyone wants to win, but managers are held to different standards than followers when they lose, especially when they lose after trying something that few others are doing.

Wayne Stewart, an associate professor of management at Clemson University, said his own research backs up the idea that owners and managers in general have different approaches to risk. While owners tend to be entrepreneurial and focused on outcomes, he said, managers are often principally focused on not screwing up.

Stewart said this might explain why coaches’ approach to risk diverges from that of owners and fans, who are principally interested in outcomes. Stewart said successful managers understand that the fear of failure is itself often the principal cause of failure.

If all economists were laid end to end, they would not reach a conclusion.” This quip from George Bernard Shaw is revealing. Economists as a group are often criticized for giving conflicting advice to policymakers. President Ronald Reagan once joked that if the game Trivial Pursuit were designed for economists, it would have 100 questions and 3,000 answers.

Why do economists so often appear to give conflicting advice to policymakers? There are two basic reasons:

- Economists may disagree about the validity of alternative positive theories about how the world works.
- Economists may have different values and therefore different normative views about what policy should try to accomplish.

Let’s discuss each of these reasons.

**Differences in Scientific Judgments**

Several centuries ago, astronomers debated whether the earth or the sun was at the center of the solar system. More recently, meteorologists have debated whether the earth is experiencing global warming and, if so, why. Science is a search for understanding about the world around us. It is not surprising that as the search continues, scientists can disagree about the direction in which truth lies.

Economists often disagree for the same reason. Economics is a young science, and there is still much to be learned. Economists sometimes disagree because they have different hunches about the validity of alternative theories or about the size of important parameters that measure how economic variables are related.

For example, economists disagree about whether the government should tax a household’s income or its consumption (spending). Advocates of a switch from the current income tax to a consumption tax believe that the change would encourage households to save more because income that is saved would not be taxed. Higher saving, in turn, would free resources for capital accumulation, leading to more rapid growth in productivity and living standards. Advocates of the current income tax system believe that household saving would not respond much to a change in the tax laws. These two groups of economists hold different normative views about the tax system because they have different positive views about the responsiveness of saving to tax incentives.

**Differences in Values**

Suppose that Peter and Paula both take the same amount of water from the town well. To pay for maintaining the well, the town taxes its residents. Peter has income of $50,000 and is taxed $5,000, or 10 percent of his income. Paula has income of $10,000 and is taxed $2,000, or 20 percent of her income.

Is this policy fair? If not, who pays too much and who pays too little? Does it matter whether Paula’s low income is due to a medical disability or to her decision...
to pursue a career in acting? Does it matter whether Peter’s high income is due to a large inheritance or to his willingness to work long hours at a dreary job?

These are difficult questions on which people are likely to disagree. If the town hired two experts to study how the town should tax its residents to pay for the well, we would not be surprised if they offered conflicting advice.

This simple example shows why economists sometimes disagree about public policy. As we learned earlier in our discussion of normative and positive analysis, policies cannot be judged on scientific grounds alone. Economists give conflicting advice sometimes because they have different values. Perfecting the science of economics will not tell us whether Peter or Paula pays too much.

**Perception versus Reality**

Because of differences in scientific judgments and differences in values, some disagreement among economists is inevitable. Yet one should not overstate the amount of disagreement. Economists agree with one another far more than is sometimes understood.

Table 1 contains 14 propositions about economic policy. In surveys of professional economists, these propositions were endorsed by an overwhelming majority of respondents. Most of these propositions would fail to command a similar consensus among the public.

**Table 1**

<table>
<thead>
<tr>
<th>Proposition (and percentage of economists who agree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A ceiling on rents reduces the quantity and quality of housing available. (93%)</td>
</tr>
<tr>
<td>2. Tariffs and import quotas usually reduce general economic welfare. (93%)</td>
</tr>
<tr>
<td>3. Flexible and floating exchange rates offer an effective international monetary arrangement. (90%)</td>
</tr>
<tr>
<td>4. Fiscal policy (e.g., tax cut and/or government expenditure increase) has a significant stimulative impact on a less than fully employed economy. (90%)</td>
</tr>
<tr>
<td>5. The United States should not restrict employers from outsourcing work to foreign countries. (90%)</td>
</tr>
<tr>
<td>6. The United States should eliminate agricultural subsidies. (85%)</td>
</tr>
<tr>
<td>7. Local and state governments should eliminate subsidies to professional sports franchises. (85%)</td>
</tr>
<tr>
<td>8. If the federal budget is to be balanced, it should be done over the business cycle rather than yearly. (85%)</td>
</tr>
<tr>
<td>9. The gap between Social Security funds and expenditures will become unsustainably large within the next 50 years if current policies remain unchanged. (85%)</td>
</tr>
<tr>
<td>10. Cash payments increase the welfare of recipients to a greater degree than do transfers-in-kind of equal cash value. (84%)</td>
</tr>
<tr>
<td>11. A large federal budget deficit has an adverse effect on the economy. (83%)</td>
</tr>
<tr>
<td>12. A minimum wage increases unemployment among young and unskilled workers. (79%)</td>
</tr>
<tr>
<td>13. The government should restructure the welfare system along the lines of a “negative income tax.” (79%)</td>
</tr>
<tr>
<td>14. Effluent taxes and marketable pollution permits represent a better approach to pollution control than imposition of pollution ceilings. (78%)</td>
</tr>
</tbody>
</table>

The first proposition in the table is about rent control, a policy that sets a legal maximum on the amount landlords can charge for their apartments. Almost all economists believe that rent control adversely affects the availability and quality of housing and is a costly way of helping the neediest members of society. Nonetheless, many city governments ignore the advice of economists and place ceilings on the rents that landlords may charge their tenants.

The second proposition in the table concerns tariffs and import quotas, two policies that restrict trade among nations. For reasons we discuss more fully later in this text, almost all economists oppose such barriers to free trade. Nonetheless, over the years, presidents and Congress have chosen to restrict the import of certain goods.

Why do policies such as rent control and trade barriers persist if the experts are united in their opposition? It may be that the realities of the political process stand as immovable obstacles. But it also may be that economists have not yet convinced enough of the public that these policies are undesirable. One purpose of this book is to help you understand the economist’s view of these and other subjects and, perhaps, to persuade you that it is the right one.

**Quick Quiz**  Why might economic advisers to the president disagree about a question of policy?

---

**LET’S GET GOING**

The first two chapters of this book have introduced you to the ideas and methods of economics. We are now ready to get to work. In the next chapter, we start learning in more detail the principles of economic behavior and economic policy.

As you proceed through this book, you will be asked to draw on many of your intellectual skills. You might find it helpful to keep in mind some advice from the great economist John Maynard Keynes:

The study of economics does not seem to require any specialized gifts of an unusually high order. Is it not . . . a very easy subject compared with the higher branches of philosophy or pure science? An easy subject, at which very few excel! The paradox finds its explanation, perhaps, in that the master-economist must possess a rare combination of gifts. He must be mathematician, historian, statesman, philosopher—in some degree. He must understand symbols and speak in words. He must contemplate the particular in terms of the general, and touch abstract and concrete in the same flight of thought. He must study the present in the light of the past for the purposes of the future. No part of man’s nature or his institutions must lie entirely outside his regard. He must be purposeful and disinterested in a simultaneous mood; as aloof and incorruptible as an artist, yet sometimes as near the earth as a politician.

It is a tall order. But with practice, you will become more and more accustomed to thinking like an economist.
Environmental Economics

Some economists are helping to save the planet.

Green Groups See Potent Tool in Economics
By Jessica E. Vascellaro

Many economists dream of getting high-paying jobs on Wall Street, at prestigious think tanks and universities or at powerful government agencies like the Federal Reserve.

But a growing number are choosing to use their skills not to track inflation or interest rates but to rescue rivers and trees. These are the “green economists,” more formally known as environmental economists, who use economic arguments and systems to persuade companies to clean up pollution and to help conserve natural areas.

Working at dozens of advocacy groups and a myriad of state and federal environmental agencies, they are helping to formulate the intellectual framework behind approaches to protecting endangered species, reducing pollution and preventing climate change. They also are becoming a link between left-leaning advocacy groups and the public and private sectors.

“In the past, many advocacy groups interpreted economics as how to make a profit or maximize income,” says Lawrence Gooder, a professor of environmental and resource economics at Stanford University in Stanford, Calif. “More economists are realizing that it offers a framework for resource allocation where resources are not only labor and capital but natural resources as well.”

Environmental economists are on the payroll of government agencies (the Environmental Protection Agency had about 164 on staff in 2004, up 36% from 1995) and groups like the Wilderness Society, a Washington-based conservation group, which has four of them to work on projects such as assessing the economic impact of building off-road driving trails. Environmental Defense, also based in Washington, was one of the first environmental-advocacy groups to hire economists and now has about eight, who do such things as develop market incentives to address environmental problems like climate change and water shortages. . .

“There used to be this idea that we shouldn’t have to monetize the environment because it is invaluable,” says Caroline Alkire, who in 1991 joined the Wilderness Society, an advocacy group in Washington, D.C., as one of the group’s first economists. “But if we are going to engage in debate on the Hill about drilling in the Arctic we need to be able to combat the financial arguments. We have to play that card or we are going to lose.”

The field of environmental economics began to take form in the 1960s when academics started to apply the tools of economics to the nascent green movement. The discipline grew more popular throughout the 1980s when the Environmental Protection Agency adopted a system of tradable permits for phasing out leaded gasoline. It wasn’t until the 1990 amendment to the Clean Air Act, however, that most environmentalists started to take economics seriously.

The amendment implemented a system of tradable allowances for acid rain, a program pushed by Environmental Defense. Under the law, plants that can reduce their emissions more cost-effectively may sell their allowances to more heavy polluters. Today, the program has exceeded its goal of reducing the amount of acid rain to half its 1980 level and is celebrated as evidence that markets can help achieve environmental goals.

Its success has convinced its former critics, who at the time contended that environmental regulation was a matter of ethics, not economics, and favored installing expensive acid rain removal technology in all power plants instead.

Greenpeace, the international environmental giant, was one of the leading opponents of the 1990 amendment. But Kert Davies, research director for Greenpeace USA, said its success and the lack of any significant action on climate policy throughout the early 1990s brought the organization around to the concept. “We now believe that [tradable permits] are the most straightforward system of reducing emissions and creating the incentives necessary for massive reductions.”

SUMMARY

- Economists try to address their subject with a scientist’s objectivity. Like all scientists, they make appropriate assumptions and build simplified models to understand the world around them. Two simple economic models are the circular-flow diagram and the production possibilities frontier.

- The field of economics is divided into two subfields: microeconomics and macroeconomics. Microeconomists study decision making by households and firms and the interaction among households and firms in the marketplace. Macroeconomists study the forces and trends that affect the economy as a whole.

- A positive statement is an assertion about how the world is. A normative statement is an assertion about how the world ought to be. When economists make normative statements, they are acting more as policy advisers than as scientists.

- Economists who advise policymakers offer conflicting advice either because of differences in scientific judgments or because of differences in values. At other times, economists are united in the advice they offer, but policymakers may choose to ignore it.

KEY CONCEPTS

circular-flow diagram, p. 24  
production possibilities frontier,  
$p. 26$  
microeconomics, p. 28  
macroeconomics, p. 28  
normative statements, p. 30  
positive statements, p. 30

QUESTIONS FOR REVIEW

1. How is economics like a science?
2. Why do economists make assumptions?
3. Should an economic model describe reality exactly?
4. Name a way that your family interacts in the factor market, and a way that it interacts in the product market.
5. Name one economic interaction that isn’t covered by the simplified circular-flow diagram.
6. Draw and explain a production possibilities frontier for an economy that produces milk and cookies. What happens to this frontier if disease kills half of the economy’s cows?
7. Use a production possibilities frontier to describe the idea of “efficiency.”
8. What are the two subfields into which economics is divided? Explain what each subfield studies.
9. What is the difference between a positive and a normative statement? Give an example of each.
10. Why do economists sometimes offer conflicting advice to policymakers?

PROBLEMS AND APPLICATIONS

1. Draw a circular-flow diagram. Identify the parts of the model that correspond to the flow of goods and services and the flow of dollars for each of the following activities.

   a. Selena pays a storekeeper $1 for a quart of milk.
   b. Stuart earns $4.50 per hour working at a fast-food restaurant.
c. Shanna spends $30 to get a haircut.
d. Sally earns $10,000 from her 10 percent ownership of Acme Industrial.

2. Imagine a society that produces military goods and consumer goods, which we’ll call “guns” and “butter.”
   a. Draw a production possibilities frontier for guns and butter. Using the concept of opportunity cost, explain why it most likely has a bowed-out shape.
   b. Show a point that is impossible for the economy to achieve. Show a point that is feasible but inefficient.
   c. Imagine that the society has two political parties, called the Hawks (who want a strong military) and the Doves (who want a smaller military). Show a point on your production possibilities frontier that the Hawks might choose and a point the Doves might choose.
   d. Imagine that an aggressive neighboring country reduces the size of its military. As a result, both the Hawks and the Doves reduce their desired production of guns by the same amount. Which party would get the bigger “peace dividend,” measured by the increase in butter production? Explain.

3. The first principle of economics discussed in Chapter 1 is that people face trade-offs. Use a production possibilities frontier to illustrate society’s trade-off between two “goods”—a clean environment and the quantity of industrial output. What do you suppose determines the shape and position of the frontier? Show what happens to the frontier if engineers develop a new way of producing electricity that emits fewer pollutants.

4. An economy consists of three workers: Larry, Moe, and Curly. Each works ten hours a day and can produce two services: mowing lawns and washing cars. In an hour, Larry can either mow one lawn or wash one car; Moe can either mow one lawn or wash two cars; and Curly can either mow two lawns or wash one car.
   a. Calculate how much of each service is produced under the following circumstances, which we label A, B, C, and D:
      • All three spend all their time mowing lawns. (A)
      • All three spend all their time washing cars. (B)
      • All three spend half their time on each activity. (C)
      • Larry spends half his time on each activity, while Moe only washes cars and Curly only mows lawns. (D)
   b. Graph the production possibilities frontier for this economy. Using your answers to part (a), identify points A, B, C, and D on your graph.
   c. Explain why the production possibilities frontier has the shape it does.
   d. Are any of the allocations calculated in part (a) inefficient? Explain.

5. Classify the following topics as relating to microeconomics or macroeconomics.
   a. a family’s decision about how much income to save
   b. the effect of government regulations on auto emissions
   c. the impact of higher national saving on economic growth
   d. a firm’s decision about how many workers to hire
   e. the relationship between the inflation rate and changes in the quantity of money

6. Classify each of the following statements as positive or normative. Explain.
   a. Society faces a short-run trade-off between inflation and unemployment.
   b. A reduction in the rate of money growth will reduce the rate of inflation.
   c. The Federal Reserve should reduce the rate of money growth.
   d. Society ought to require welfare recipients to look for jobs.
   e. Lower tax rates encourage more work and more saving.

7. Classify each of the statements in Table 1 as positive, normative, or ambiguous. Explain.

8. If you were president, would you be more interested in your economic advisers’ positive views or their normative views? Why?

9. Find a recent copy of the Economic Report of the President at your library or on the Internet (http://www.gpoaccess.gov/eop/index.html). Read a chapter about an issue that interests you. Summarize the economic problem at hand and describe the council’s recommended policy.
**GRAPHING: A BRIEF REVIEW**

Many of the concepts that economists study can be expressed with numbers—the price of bananas, the quantity of bananas sold, the cost of growing bananas, and so on. Often, these economic variables are related to one another. When the price of bananas rises, people buy fewer bananas. One way of expressing the relationships among variables is with graphs.

Graphs serve two purposes. First, when developing economic theories, graphs offer a way to visually express ideas that might be less clear if described with equations or words. Second, when analyzing economic data, graphs provide a powerful way of finding and interpreting patterns. Whether we are working with theory or with data, graphs provide a lens through which a recognizable forest emerges from a multitude of trees.

Numerical information can be expressed graphically in many ways, just as there are many ways to express a thought in words. A good writer chooses words that will make an argument clear, a description pleasing, or a scene dramatic. An effective economist chooses the type of graph that best suits the purpose at hand.

In this appendix, we discuss how economists use graphs to study the mathematical relationships among variables. We also discuss some of the pitfalls that can arise in the use of graphical methods.

**Graphs of a Single Variable**

Three common graphs are shown in Figure A-1. The pie chart in panel (a) shows how U.S. national income is derived from various sources. The bar graph in panel (b) compares the average income in four countries. The time-series graph in panel (c) shows the productivity of labor in U.S. businesses from 1950 to 2000.
income for four countries. The height of each bar represents the average income in each country. The time-series graph in panel (c) traces the rising productivity in the U.S. business sector over time. The height of the line shows output per hour in each year. You have probably seen similar graphs in newspapers and magazines.

**Graphs of Two Variables: The Coordinate System**

Although the three graphs in Figure A-1 are useful in showing how a variable changes over time or across individuals, such graphs are limited in how much they can tell us. These graphs display information only on a single variable. Economists are often concerned with the relationships between variables. Thus, they need to display two variables on a single graph. The **coordinate system** makes this possible.

Suppose you want to examine the relationship between study time and grade point average. For each student in your class, you could record a pair of numbers: hours per week spent studying and grade point average. These numbers could then be placed in parentheses as an *ordered pair* and appear as a single point on the graph. Albert E., for instance, is represented by the ordered pair (25 hours/week, 3.5 GPA), while his “what-me-worry?” classmate Alfred E. is represented by the ordered pair (5 hours/week, 2.0 GPA).

We can graph these ordered pairs on a two-dimensional grid. The first number in each ordered pair, called the *x-coordinate*, tells us the horizontal location of the point. The second number, called the *y-coordinate*, tells us the vertical location of the point. The point with both an x-coordinate and a y-coordinate of zero is known as the **origin**. The two coordinates in the ordered pair tell us where the point is located in relation to the origin: x units to the right of the origin and y units above it.

Figure A-2 graphs grade point average against study time for Albert E., Alfred E., and their classmates. This type of graph is called a *scatterplot* because it plots scattered points. Looking at this graph, we immediately notice that points farther to the right (indicating more study time) also tend to be higher (indicating a better

**Figure A-2**

*Using the Coordinate System*

Grade point average is measured on the vertical axis and study time on the horizontal axis. Albert E., Alfred E., and their classmates are represented by various points. We can see from the graph that students who study more tend to get higher grades.
grade point average). Because study time and grade point average typically move in the same direction, we say that these two variables have a *positive correlation*. By contrast, if we were to graph party time and grades, we would likely find that higher party time is associated with lower grades; because these variables typically move in opposite directions, we call this a *negative correlation*. In either case, the coordinate system makes the correlation between the two variables easy to see.

**Curves in the Coordinate System**

Students who study more do tend to get higher grades, but other factors also influence a student’s grade. Previous preparation is an important factor, for instance, as are talent, attention from teachers, even eating a good breakfast. A scatterplot like Figure A-2 does not attempt to isolate the effect that study has on grades from the effects of other variables. Often, however, economists prefer looking at how one variable affects another, holding everything else constant.

To see how this is done, let’s consider one of the most important graphs in economics: the demand curve. The demand curve traces out the effect of a good’s price on the quantity of the good consumers want to buy. Before showing a demand curve, however, consider Table A-1, which shows how the number of novels that Emma buys depends on her income and on the price of novels. When novels are cheap, Emma buys them in large quantities. As they become more expensive, she instead borrows books from the library or chooses to go to the movies rather than read. Similarly, at any given price, Emma buys more novels when she has a higher income. That is, when her income increases, she spends part of the additional income on novels and part on other goods.

We now have three variables—the price of novels, income, and the number of novels purchased—which are more than we can represent in two dimensions. To put the information from Table A-1 in graphical form, we need to hold one of the three variables constant and trace out the relationship between the other two. Because the demand curve represents the relationship between price and quantity demanded, we hold Emma’s income constant and show how the number of novels she buys varies with the price of novels.

<table>
<thead>
<tr>
<th>A-1 TABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Novels Purchased by Emma</strong></td>
</tr>
<tr>
<td>This table shows the number of novels Emma buys at various incomes and prices. For any given level of income, the data on price and quantity demanded can be graphed to produce Emma’s demand curve for novels, as shown in Figures A-3 and A-4.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Price</th>
<th>$20,000</th>
<th>$30,000</th>
<th>$40,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10</td>
<td>2 novels</td>
<td>5 novels</td>
<td>8 novels</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>21</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>Demand curve, $D_3$</td>
<td>Demand curve, $D_1$</td>
<td>Demand curve, $D_2$</td>
<td></td>
</tr>
</tbody>
</table>
Suppose that Emma’s income is $30,000 per year. If we place the number of novels Emma purchases on the x-axis and the price of novels on the y-axis, we can graphically represent the middle column of Table A-1. When the points that represent these entries from the table—(5 novels, $10), (9 novels, $9), and so on—are connected, they form a line. This line, pictured in Figure A-3, is known as Emma’s demand curve for novels; it tells us how many novels Emma purchases at any given price. The demand curve is downward sloping, indicating that a higher price reduces the quantity of novels demanded. Because the quantity of novels demanded and the price move in opposite directions, we say that the two variables are negatively related. (Conversely, when two variables move in the same direction, the curve relating them is upward sloping, and we say the variables are positively related.)

Now suppose that Emma’s income rises to $40,000 per year. At any given price, Emma will purchase more novels than she did at her previous level of income. Just as earlier we drew Emma’s demand curve for novels using the entries from the middle column of Table A-1, we now draw a new demand curve using the entries from the right column of the table. This new demand curve (curve D₂) is pictured alongside the old one (curve D₁) in Figure A-4; the new curve is a similar line drawn farther to the right. We therefore say that Emma’s demand curve for novels shifts to the right when her income increases. Likewise, if Emma’s income were to fall to $20,000 per year, she would buy fewer novels at any given price and her demand curve would shift to the left (to curve D₃).

In economics, it is important to distinguish between movements along a curve and shifts of a curve. As we can see from Figure A-3, if Emma earns $30,000 per year and novels cost $8 apiece, she will purchase 13 novels per year. If the price of...
novels falls to $7, Emma will increase her purchases of novels to 17 per year. The demand curve, however, stays fixed in the same place. Emma still buys the same number of novels at each price, but as the price falls, she moves along her demand curve from left to right. By contrast, if the price of novels remains fixed at $8 but her income rises to $40,000, Emma increases her purchases of novels from 13 to 16 per year. Because Emma buys more novels at each price, her demand curve shifts out, as shown in Figure A-4.

There is a simple way to tell when it is necessary to shift a curve: *When a variable that is not named on either axis changes, the curve shifts.* Income is on neither the x-axis nor the y-axis of the graph, so when Emma’s income changes, her demand curve must shift. The same is true for any change that affects Emma’s purchasing habits besides a change in the price of novels. If, for instance, the public library closes and Emma must buy all the books she wants to read, she will demand more novels at each price, and her demand curve will shift to the right. Or if the price of movies falls and Emma spends more time at the movies and less time reading, she will demand fewer novels at each price, and her demand curve will shift to the left. By contrast, when a variable on an axis of the graph changes, the curve does not shift. We read the change as a movement along the curve.

**Slope**

One question we might want to ask about Emma is how much her purchasing habits respond to price. Look at the demand curve pictured in Figure A-5. If this curve is very steep, Emma purchases nearly the same number of novels regardless of whether they are cheap or expensive. If this curve is much flatter, Emma
purchases many fewer novels when the price rises. To answer questions about how much one variable responds to changes in another variable, we can use the concept of slope.

The slope of a line is the ratio of the vertical distance covered to the horizontal distance covered as we move along the line. This definition is usually written out in mathematical symbols as follows:

$$ \text{slope} = \frac{\Delta y}{\Delta x} $$

where the Greek letter $\Delta$ (delta) stands for the change in a variable. In other words, the slope of a line is equal to the “rise” (change in $y$) divided by the “run” (change in $x$). The slope will be a small positive number for a fairly flat upward-sloping line, a large positive number for a steep upward-sloping line, and a negative number for a downward-sloping line. A horizontal line has a slope of zero because in this case the $y$-variable never changes; a vertical line is said to have an infinite slope because the $y$-variable can take any value without the $x$-variable changing at all.

What is the slope of Emma’s demand curve for novels? First of all, because the curve slopes down, we know the slope will be negative. To calculate a numerical value for the slope, we must choose two points on the line. With Emma’s income at $30,000, she will purchase 21 novels at a price of $6 or 13 novels at a price of $8. When we apply the slope formula, we are concerned with the change between these two points; in other words, we are concerned with the difference between

![Diagram](Figure A-5)

**Calculating the Slope of a Line**
To calculate the slope of the demand curve, we can look at the changes in the $x$- and $y$-coordinates as we move from the point (21 novels, $6) to the point (13 novels, $8). The slope of the line is the ratio of the change in the $y$-coordinate ($-2$) to the change in the $x$-coordinate ($+8$), which equals $-\frac{1}{4}$. 
them, which lets us know that we will have to subtract one set of values from the other, as follows:

\[
\text{slope} = \frac{\Delta y}{\Delta x} = \frac{\text{first } y\text{-coordinate} - \text{second } y\text{-coordinate}}{\text{first } x\text{-coordinate} - \text{second } x\text{-coordinate}} = \frac{6 - 8}{21 - 13} = \frac{-2}{8} = -\frac{1}{4}
\]

Figure A-5 shows graphically how this calculation works. Try computing the slope of Emma’s demand curve using two different points. You should get exactly the same result, \(-\frac{1}{4}\). One of the properties of a straight line is that it has the same slope everywhere. This is not true of other types of curves, which are steeper in some places than in others.

The slope of Emma’s demand curve tells us something about how responsive her purchases are to changes in the price. A small slope (a number close to zero) means that Emma’s demand curve is relatively flat; in this case, she adjusts the number of novels she buys substantially in response to a price change. A larger slope (a number farther from zero) means that Emma’s demand curve is relatively steep; in this case, she adjusts the number of novels she buys only slightly in response to a price change.

**Cause and Effect**

Economists often use graphs to advance an argument about how the economy works. In other words, they use graphs to argue about how one set of events causes another set of events. With a graph like the demand curve, there is no doubt about cause and effect. Because we are varying price and holding all other variables constant, we know that changes in the price of novels cause changes in the quantity Emma demands. Remember, however, that our demand curve came from a hypothetical example. When graphing data from the real world, it is often more difficult to establish how one variable affects another.

The first problem is that it is difficult to hold everything else constant when studying the relationship between two variables. If we are not able to hold other variables constant, we might decide that one variable on our graph is causing changes in the other variable when actually those changes are caused by a third omitted variable not pictured on the graph. Even if we have identified the correct two variables to look at, we might run into a second problem—reverse causality. In other words, we might decide that A causes B when in fact B causes A. The omitted-variable and reverse-causality traps require us to proceed with caution when using graphs to draw conclusions about causes and effects.

**Omitted Variables**

To see how omitting a variable can lead to a deceptive graph, let’s consider an example. Imagine that the government, spurred by public concern about the large number of deaths from cancer, commissions an exhaustive study from Big Brother Statistical Services, Inc. Big Brother examines many of the items found in people’s homes to see which of them are associated with the risk of cancer. Big Brother reports a strong relationship between two variables: the number of cigarette lighters that a household owns and the probability that someone in the household will develop cancer. Figure A-6 shows this relationship.

What should we make of this result? Big Brother advises a quick policy response. It recommends that the government discourage the ownership of cigarette lighters by taxing their sale. It also recommends that the government require
warning labels: “Big Brother has determined that this lighter is dangerous to your health.”

In judging the validity of Big Brother’s analysis, one question is paramount: Has Big Brother held constant every relevant variable except the one under consideration? If the answer is no, the results are suspect. An easy explanation for Figure A-6 is that people who own more cigarette lighters are more likely to smoke cigarettes and that cigarettes, not lighters, cause cancer. If Figure A-6 does not hold constant the amount of smoking, it does not tell us the true effect of owning a cigarette lighter.

This story illustrates an important principle: When you see a graph used to support an argument about cause and effect, it is important to ask whether the movements of an omitted variable could explain the results you see.

**Reverse Causality** Economists can also make mistakes about causality by misreading its direction. To see how this is possible, suppose the Association of American Anarchists commissions a study of crime in America and arrives at Figure A-7, which plots the number of violent crimes per thousand people in major cities against the number of police officers per thousand people. The anarchists note the curve’s upward slope and argue that because police increase rather than decrease the amount of urban violence, law enforcement should be abolished.

If we could run a controlled experiment, we would avoid the danger of reverse causality. To run an experiment, we would set the number of police officers in different cities randomly and then examine the correlation between police and crime. Figure A-7, however, is not based on such an experiment. We simply observe that more dangerous cities have more police officers. The explanation for this may be that more dangerous cities hire more police. In other words, rather than police causing crime, crime may cause police. Nothing in the graph itself allows us to establish the direction of causality.

It might seem that an easy way to determine the direction of causality is to examine which variable moves first. If we see crime increase and then the police force expand, we reach one conclusion. If we see the police force expand and then crime increase, we reach the other. Yet there is also a flaw with this approach: Often, people change their behavior not in response to a change in their present conditions but in response to a change in their expectations of future conditions. A

---

**Figure A-6**

*Graph with an Omitted Variable*

The upward-sloping curve shows that members of households with more cigarette lighters are more likely to develop cancer. Yet we should not conclude that ownership of lighters causes cancer because the graph does not take into account the number of cigarettes smoked.
city that expects a major crime wave in the future, for instance, might hire more police now. This problem is even easier to see in the case of babies and minivans. Couples often buy a minivan in anticipation of the birth of a child. The minivan comes before the baby, but we wouldn’t want to conclude that the sale of minivans causes the population to grow!

There is no complete set of rules that says when it is appropriate to draw causal conclusions from graphs. Yet just keeping in mind that cigarette lighters don’t cause cancer (omitted variable) and minivans don’t cause larger families (reverse causality) will keep you from falling for many faulty economic arguments.
Interdependence and the Gains from Trade

Consider your typical day. You wake up in the morning and pour yourself juice from oranges grown in Florida and coffee from beans grown in Brazil. Over breakfast, you watch a news program broadcast from New York on your television made in Japan. You get dressed in clothes made of cotton grown in Georgia and sewn in factories in Thailand. You drive to class in a car made of parts manufactured in more than a dozen countries around the world. Then you open up your economics textbook written by an author living in Massachusetts, published by a company located in Ohio, and printed on paper made from trees grown in Oregon.

Every day, you rely on many people, most of whom you have never met, to provide you with the goods and services that you enjoy. Such interdependence is possible because people trade with one another. Those people providing you goods and services are not acting out of generosity. Nor is some government agency directing them to satisfy your desires. Instead, people provide you and other consumers with the goods and services they produce because they get something in return.

In subsequent chapters, we examine how our economy coordinates the activities of millions of people with varying tastes and abilities. As a starting point for
this analysis, here we consider the reasons for economic interdependence. One of the Ten Principles of Economics highlighted in Chapter 1 is that trade can make everyone better off. In this chapter, we examine this principle more closely. What exactly do people gain when they trade with one another? Why do people choose to become interdependent?

The answers to these questions are key to understanding the modern global economy. In most countries today, many goods and services consumed are imported from abroad, and many goods and services produced are exported to foreign customers. The analysis in this chapter explains interdependence not only among individuals but also among nations. As we will see, the gains from trade are much the same whether you are buying a haircut from your local barber or a T-shirt made by a worker on the other side of the globe.

A PARABLE FOR THE MODERN ECONOMY

To understand why people choose to depend on others for goods and services and how this choice improves their lives, let’s look at a simple economy. Imagine that there are two goods in the world: meat and potatoes. And there are two people in the world—a cattle rancher and a potato farmer—each of whom would like to eat both meat and potatoes.

The gains from trade are most obvious if the rancher can produce only meat and the farmer can produce only potatoes. In one scenario, the rancher and the farmer could choose to have nothing to do with each other. But after several months of eating beef roasted, boiled, broiled, and grilled, the rancher might decide that self-sufficiency is not all it’s cracked up to be. The farmer, who has been eating potatoes mashed, fried, baked, and scalloped, would likely agree. It is easy to see that trade would allow them to enjoy greater variety: Each could then have a steak with a baked potato or a burger with fries.

Although this scene illustrates most simply how everyone can benefit from trade, the gains would be similar if the rancher and the farmer were each capable of producing the other good, but only at great cost. Suppose, for example, that the potato farmer is able to raise cattle and produce meat, but that he is not very good at it. Similarly, suppose that the cattle rancher is able to grow potatoes but that her land is not very well suited for it. In this case, the farmer and the rancher can each benefit by specializing in what he or she does best and then trading with the other.

The gains from trade are less obvious, however, when one person is better at producing every good. For example, suppose that the rancher is better at raising cattle and better at growing potatoes than the farmer. In this case, should the rancher choose to remain self-sufficient? Or is there still reason for her to trade with the farmer? To answer this question, we need to look more closely at the factors that affect such a decision.

PRODUCTION POSSIBILITIES

Suppose that the farmer and the rancher each work 8 hours per day and can devote this time to growing potatoes, raising cattle, or a combination of the two. The table in Figure 1 shows the amount of time each person requires to produce 1 ounce of each good. The farmer can produce an ounce of potatoes in 15 minutes
and an ounce of meat in 60 minutes. The rancher, who is more productive in both activities, can produce an ounce of potatoes in 10 minutes and an ounce of meat in 20 minutes. The last two columns in the table show the amounts of meat or potatoes the farmer and rancher can produce if they work an 8-hour day producing only that good.

Panel (b) of Figure 1 illustrates the amounts of meat and potatoes that the farmer can produce. If the farmerdevotes all 8 hours of his time to potatoes, he produces 32 ounces of potatoes (measured on the horizontal axis) and no meat. If he devotes all his time to meat, he produces 8 ounces of meat (measured on the vertical axis) and no potatoes. If the farmer divides his time equally between the two activities, spending 4 hours on each, he produces 16 ounces of potatoes and 4 ounces of meat. The figure shows these three possible outcomes and all others in between.

This graph is the farmer’s production possibilities frontier. As we discussed in Chapter 2, a production possibilities frontier shows the various mixes of output that an economy can produce. It illustrates one of the Ten Principles of Economics in Chapter 1: People face trade-offs. Here the farmer faces a trade-off between producing meat and producing potatoes.
You may recall that the production possibilities frontier in Chapter 2 was drawn bowed out. In that case, the rate at which society could trade one good for the other depended on the amounts that were being produced. Here, however, the farmer’s technology for producing meat and potatoes (as summarized in Figure 1) allows him to switch between the two goods at a constant rate. Whenever the farmer spends 1 hour less producing meat and 1 hour more producing potatoes, he reduces his output of meat by 1 ounce and raises his output of potatoes by 4 ounces—and this is true regardless of how much he is already producing. As a result, the production possibilities frontier is a straight line.

Panel (c) of Figure 1 shows the production possibilities frontier for the rancher. If the rancher devotes all 8 hours of her time to potatoes, she produces 48 ounces of potatoes and no meat. If she devotes all her time to meat, she produces 24 ounces of meat and no potatoes. If the rancher divides her time equally, spending 4 hours on each activity, she produces 24 ounces of potatoes and 12 ounces of meat. Once again, the production possibilities frontier shows all the possible outcomes.

If the farmer and rancher choose to be self-sufficient rather than trade with each other, then each consumes exactly what he or she produces. In this case, the production possibilities frontier is also the consumption possibilities frontier. That is, without trade, Figure 1 shows the possible combinations of meat and potatoes that the farmer and rancher can each produce and then consume.

These production possibilities frontiers are useful in showing the trade-offs that the farmer and rancher face, but they do not tell us what the farmer and rancher will actually choose to do. To determine their choices, we need to know the tastes of the farmer and the rancher. Let’s suppose they choose the combinations identified by points A and B in Figure 1: The farmer produces and consumes 16 ounces of potatoes and 4 ounces of meat, while the rancher produces and consumes 24 ounces of potatoes and 12 ounces of meat.

**Specialization and Trade**

After several years of eating combination B, the rancher gets an idea and goes to talk to the farmer:

**Rancher:** Farmer, my friend, have I got a deal for you! I know how to improve life for both of us. I think you should stop producing meat altogether and devote all your time to growing potatoes. According to my calculations, if you work 8 hours a day growing potatoes, you’ll produce 32 ounces of potatoes. If you give me 15 of those 32 ounces, I’ll give you 5 ounces of meat in return. In the end, you’ll get to eat 17 ounces of potatoes and 5 ounces of meat every day, instead of the 16 ounces of potatoes and 4 ounces of meat you now get. If you go along with my plan, you’ll have more of both foods. [To illustrate her point, the rancher shows the farmer panel (a) of Figure 2.]

**Farmer:** (sounding skeptical) That seems like a good deal for me. But I don’t understand why you are offering it. If the deal is so good for me, it can’t be good for you too.

**Rancher:** Oh, but it is! Suppose I spend 6 hours a day raising cattle and 2 hours growing potatoes. Then I can produce 18 ounces of meat and 12 ounces of potatoes. After I give you 5 ounces of my meat in exchange for 15 ounces of your potatoes, I’ll end up with 13 ounces of meat and 27 ounces of potatoes, instead of the 12 ounces of meat and
The proposed trade between the farmer and the rancher offers each of them a combination of meat and potatoes that would be impossible in the absence of trade. In panel (a), the farmer gets to consume at point A* rather than point A. In panel (b), the rancher gets to consume at point B* rather than point B. Trade allows each to consume more meat and more potatoes.

24 ounces of potatoes that I now get. So I will also consume more of both foods than I do now. [She points out panel (b) of Figure 2.]

Farmer: I don’t know. . . . This sounds too good to be true.

Rancher: It’s really not as complicated as it first seems. Here—I’ve summarized my proposal for you in a simple table. [The rancher shows the farmer a copy of the table at the bottom of Figure 2.]

Farmer: (after pausing to study the table) These calculations seem correct, but I am puzzled. How can this deal make us both better off?

Rancher: We can both benefit because trade allows each of us to specialize in doing what we do best. You will spend more time growing potatoes and less time raising cattle. I will spend more time raising cattle and
The rancher’s explanation of the gains from trade, though correct, poses a puzzle: If the rancher is better at both raising cattle and growing potatoes, how can the farmer ever specialize in doing what he does best? The farmer doesn’t seem to do anything best. To solve this puzzle, we need to look at the principle of comparative advantage.

As a first step in developing this principle, consider the following question: In our example, who can produce potatoes at a lower cost—the farmer or the rancher? There are two possible answers, and in these two answers lie the solution to our puzzle and the key to understanding the gains from trade.

**ABSOLUTE ADVANTAGE**

One way to answer the question about the cost of producing potatoes is to compare the inputs required by the two producers. Economists use the term **absolute advantage** when comparing the productivity of one person, firm, or nation to that of another. The producer that requires a smaller quantity of inputs to produce a good is said to have an absolute advantage in producing that good.

In our example, time is the only input, so we can determine absolute advantage by looking at how much time each type of production takes. The rancher has an absolute advantage both in producing meat and in producing potatoes because she requires less time than the farmer to produce a unit of either good. The rancher needs to input only 20 minutes to produce an ounce of meat, whereas the farmer needs 60 minutes. Similarly, the rancher needs only 10 minutes to produce an ounce of potatoes, whereas the farmer needs 15 minutes. Based on this information, we can conclude that the rancher has the lower cost of producing potatoes, if we measure cost by the quantity of inputs.

**OPPORTUNITY COST AND COMPARATIVE ADVANTAGE**

There is another way to look at the cost of producing potatoes. Rather than comparing inputs required, we can compare the opportunity costs. Recall from Chapter 1 that the **opportunity cost** of some item is what we give up to get that item. In our example, we assumed that the farmer and the rancher each spend 8 hours a day working. Time spent producing potatoes, therefore, takes away from time available for producing meat. When reallocating time between the two goods, the rancher and farmer give up units of one good to produce units of the other, thereby moving along the production possibilities frontier. The opportunity cost measures the trade-off between the two goods that each producer faces.

**QUICK QUIZ**

Draw an example of a production possibilities frontier for Robinson Crusoe, a shipwrecked sailor who spends his time gathering coconuts and catching fish. Does this frontier limit Crusoe’s consumption of coconuts and fish if he lives by himself? Does he face the same limits if he can trade with natives on the island?
Let’s first consider the rancher’s opportunity cost. According to the table in panel (a) of Figure 1, producing 1 ounce of potatoes takes 10 minutes of work. When the rancher spends those 10 minutes producing potatoes, she spends 10 minutes less producing meat. Because the rancher needs 20 minutes to produce 1 ounce of meat, 10 minutes of work would yield \(\frac{1}{2}\) ounce of meat. Hence, the rancher’s opportunity cost of producing 1 ounce of potatoes is \(\frac{1}{2}\) ounce of meat.

Now consider the farmer’s opportunity cost. Producing 1 ounce of potatoes takes him 15 minutes. Because he needs 60 minutes to produce 1 ounce of meat, 15 minutes of work would yield \(\frac{1}{4}\) ounce of meat. Hence, the farmer’s opportunity cost of 1 ounce of potatoes is \(\frac{1}{4}\) ounce of meat.

Table 1 shows the opportunity costs of meat and potatoes for the two producers. Notice that the opportunity cost of meat is the inverse of the opportunity cost of potatoes. Because 1 ounce of potatoes costs the rancher \(\frac{1}{2}\) ounce of meat, 1 ounce of meat costs the rancher 2 ounces of potatoes. Similarly, because 1 ounce of potatoes costs the farmer \(\frac{1}{4}\) ounce of meat, 1 ounce of meat costs the farmer 4 ounces of potatoes.

Economists use the term **comparative advantage** when describing the opportunity cost of two producers. The producer who gives up less of other goods to produce Good X has the smaller opportunity cost of producing Good X and is said to have a comparative advantage in producing it. In our example, the farmer has a lower opportunity cost of producing potatoes than the rancher: An ounce of potatoes costs the farmer only \(\frac{1}{4}\) ounce of meat, but it costs the rancher \(\frac{1}{2}\) ounce of meat. Conversely, the rancher has a lower opportunity cost of producing meat than the farmer: An ounce of meat costs the rancher 2 ounces of potatoes, but it costs the farmer 4 ounces of potatoes. Thus, the farmer has a comparative advantage in growing potatoes, and the rancher has a comparative advantage in producing meat.

Although it is possible for one person to have an absolute advantage in both goods (as the rancher does in our example), it is impossible for one person to have a comparative advantage in both goods. Because the opportunity cost of one good is the inverse of the opportunity cost of the other, if a person’s opportunity cost of one good is relatively high, the opportunity cost of the other good must be relatively low. Comparative advantage reflects the relative opportunity cost. Unless two people have exactly the same opportunity cost, one person will have a comparative advantage in one good, and the other person will have a comparative advantage in the other good.

**Comparative Advantage and Trade**

The gains from specialization and trade are based not on absolute advantage but on comparative advantage. When each person specializes in producing the goods

<table>
<thead>
<tr>
<th>Opportunity Cost of:</th>
<th>1 oz of Meat</th>
<th>1 oz of Potatoes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farmer</td>
<td>4 oz potatoes</td>
<td>(\frac{1}{4}) oz meat</td>
</tr>
<tr>
<td>Rancher</td>
<td>2 oz potatoes</td>
<td>(\frac{1}{2}) oz meat</td>
</tr>
</tbody>
</table>
for which he or she has a comparative advantage, total production in the economy rises. This increase in the size of the economic pie can be used to make everyone better off.

In our example, the farmer spends more time growing potatoes, and the rancher spends more time producing meat. As a result, the total production of potatoes rises from 40 to 44 ounces, and the total production of meat rises from 16 to 18 ounces. The farmer and rancher share the benefits of this increased production.

We can also look at the gains from trade in terms of the price that each party pays the other. Because the farmer and rancher have different opportunity costs, they can both get a bargain. That is, each benefits from trade by obtaining a good at a price that is lower than his or her opportunity cost of that good.

Consider the proposed deal from the viewpoint of the farmer. The farmer gets 5 ounces of meat in exchange for 15 ounces of potatoes. In other words, the farmer buys each ounce of meat for a price of 3 ounces of potatoes. This price of meat is lower than his opportunity cost for an ounce of meat, which is 4 ounces of potatoes. Thus, the farmer benefits from the deal because he gets to buy meat at a good price.

Now consider the deal from the rancher’s viewpoint. The rancher buys 15 ounces of potatoes for a price of 5 ounces of meat. That is, the price of potatoes is ¼ ounce of meat. This price of potatoes is lower than her opportunity cost of an ounce of potatoes, which is ½ ounce of meat. The rancher benefits because she gets to buy potatoes at a good price.

The moral of the story of the farmer and the rancher should now be clear: Trade can benefit everyone in society because it allows people to specialize in activities in which they have a comparative advantage.

**The Price of the Trade**

The principle of comparative advantage establishes that there are gains from specialization and trade, but it leaves open a couple of related questions: What determines the price at which trade takes place? How are the gains from trade shared between the trading parties? The precise answer to these questions is beyond the scope of this chapter, but we can state one general rule: *For both parties to gain from trade, the price at which they trade must lie between the two opportunity costs.*

In our example, the farmer and rancher agreed to trade at a rate of 3 ounces of potatoes for each ounce of meat. This price is between the rancher’s opportunity cost (2 ounces of potatoes per ounce of meat) and the farmer’s opportunity cost (4 ounces of potatoes per ounce of meat). The price need not be exactly in the middle for both parties to gain, but it must be somewhere between 2 and 4.

To see why the price has to be in this range, consider what would happen if it were not. If the price of meat were below 2 ounces of potatoes, both the farmer and the rancher would want to buy meat, because the price would be below their opportunity costs. Similarly, if the price of meat were above 4 ounces of potatoes, both would want to sell meat, because the price would be above their opportunity costs. But there are only two members of this economy. They cannot both be buyers of meat, nor can they both be sellers. Someone has to take the other side of the deal.

A mutually advantageous trade can be struck at a price between 2 and 4. In this price range, the rancher wants to sell meat to buy potatoes, and the farmer wants to sell potatoes to buy meat. Each party can buy a good at a price that is lower
than his or her opportunity cost. In the end, both of them specialize in the good for which he or she has a comparative advantage and are, as a result, better off.

**QUICK QUIZ**  
Robinson Crusoe can gather 10 coconuts or catch 1 fish per hour. His friend Friday can gather 30 coconuts or catch 2 fish per hour. What is Crusoe’s opportunity cost of catching one fish? What is Friday’s? Who has an absolute advantage in catching fish? Who has a comparative advantage in catching fish?

**APPLICATIONS OF COMPARATIVE ADVANTAGE**  
The principle of comparative advantage explains interdependence and the gains from trade. Because interdependence is so prevalent in the modern world, the principle of comparative advantage has many applications. Here are two examples, one fanciful and one of great practical importance.
SHOULD TIGER WOODS MOW HIS OWN LAWN?

Tiger Woods spends a lot of time walking around on grass. One of the most talented golfers of all time, he can hit a drive and sink a putt in a way that most casual golfers only dream of doing. Most likely, he is talented at other activities too. For example, let’s imagine that Woods can mow his lawn faster than anyone else. But just because he can mow his lawn fast, does this mean he should?

To answer this question, we can use the concepts of opportunity cost and comparative advantage. Let’s say that Woods can mow his lawn in 2 hours. In that same 2 hours, he could film a television commercial for Nike and earn $10,000. By contrast, Forrest Gump, the boy next door, can mow Woods’s lawn in 4 hours. In that same 4 hours, he could work at McDonald’s and earn $20.

In this example, Woods has an absolute advantage in mowing lawns because he can do the work with a lower input of time. Yet because Woods’s opportunity cost of mowing the lawn is $10,000 and Forrest’s opportunity cost is only $20, Forrest has a comparative advantage in mowing lawns.

The gains from trade in this example are tremendous. Rather than mowing his own lawn, Woods should make the commercial and hire Forrest to mow the lawn. As long as Woods pays Forrest more than $20 and less than $10,000, both of them are better off.

SHOULD THE UNITED STATES TRADE WITH OTHER COUNTRIES?

Just as individuals can benefit from specialization and trade with one another, as the farmer and rancher did, so can populations of people in different countries. Many of the goods that Americans enjoy are produced abroad, and many of the goods produced in the United States are sold abroad. Goods produced abroad and sold domestically are called imports. Goods produced domestically and sold abroad are called exports.

To see how countries can benefit from trade, suppose there are two countries, the United States and Japan, and two goods, food and cars. Imagine that the two countries produce cars equally well: An American worker and a Japanese worker can each produce one car per month. By contrast, because the United States has more and better land, it is better at producing food: A U.S. worker can produce 2 tons of food per month, whereas a Japanese worker can produce only 1 ton of food per month.

The principle of comparative advantage states that each good should be produced by the country that has the smaller opportunity cost of producing that good. Because the opportunity cost of a car is 2 tons of food in the United States but only 1 ton of food in Japan, Japan has a comparative advantage in producing cars. Japan should produce more cars than it wants for its own use and export some of them to the United States. Similarly, because the opportunity cost of a ton of food is 1 car in Japan but only ½ car in the United States, the United States has a comparative advantage in producing food. The United States should produce more food than it wants to consume and export some to Japan. Through specialization and trade, both countries can have more food and more cars.

In reality, of course, the issues involved in trade among nations are more complex than this example suggests. Most important among these issues is that each country has many citizens with different interests. International trade can make some individuals worse off, even as it makes the country as a whole better off.
When the United States exports food and imports cars, the impact on an American farmer is not the same as the impact on an American autoworker. Yet, contrary to the opinions sometimes voiced by politicians and pundits, international trade is not like war, in which some countries win and others lose. Trade allows all countries to achieve greater prosperity.

**Quick Quiz** Suppose that a skilled brain surgeon also happens to be the world’s fastest typist. Should she do her own typing or hire a secretary? Explain.

---

**In The News**

**The Changing Face of International Trade**

A decade ago, no one would have asked which nation has a comparative advantage in slaying ogres. But technology is rapidly changing the goods and services that are traded across national borders.

**Ogre to Slay? Outsource It to Chinese**

By David Barboza

Fuzhou, China—One of China’s newest factories operates here in the basement of an old warehouse. Posters of World of Warcraft and Magic Land hang above a corps of young people glued to their computer screens, pounding away at their keyboards in the latest hustle for money.

The people working at this clandestine locale are “gold farmers.” Every day, in 12-hour shifts, they “play” computer games by killing onscreen monsters and winning battles, harvesting artificial gold coins and other virtual goods as rewards that, as it turns out, can be transformed into real cash.

That is because, from Seoul to San Francisco, affluent online gamers who lack the time and patience to work their way up to the higher levels of gamedom are willing to pay the young Chinese here to play the early rounds for them.

“For 12 hours a day, 7 days a week, my colleagues and I are killing monsters,” said a 23-year-old gamer who works here in this makeshift factory and goes by the online code name Wandering. “I make about $250 a month, which is pretty good compared with the other jobs I’ve had. And I can play games all day.”

He and his comrades have created yet another new business out of cheap Chinese labor. They are tapping into the fast-growing world of “massively multiplayer online games,” which involve role playing and often revolve around fantasy or warfare in medieval kingdoms or distant galaxies.

For the Chinese in game-playing factories like these, though, it is not all fun and games. These workers have strict quotas and are supervised by bosses who equip them with computers, software and Internet connections to thrash online trolls, gnomes and ogres.

As they grind through the games, they accumulate virtual currency that is valuable to game players around the world. The games allow players to trade currency to other players, who can then use it to buy better armor, amulets, magic spells and other accoutrements to climb to higher levels or create more powerful characters.

The Internet is now filled with classified advertisements from small companies—many of them here in China—auctioning for real money their powerful figures, called avatars.

“It’s unimaginable how big this is,” says Chen Yu, 27, who employs 20 full-time gamers here in Fuzhou. “They say that in some of these popular games, 40 or 50 percent of the players are actually Chinese farmers.”

You should now understand more fully the benefits of living in an interdependent economy. When Americans buy tube socks from China, when residents of Maine drink orange juice from Florida, and when a homeowner hires the kid next door to mow the lawn, the same economic forces are at work. The principle of comparative advantage shows that trade can make everyone better off.

Having seen why interdependence is desirable, you might naturally ask how it is possible. How do free societies coordinate the diverse activities of all the people involved in their economies? What ensures that goods and services will get from those who should be producing them to those who should be consuming them? In a world with only two people, such as the rancher and the farmer, the answer is simple: These two people can bargain and allocate resources between themselves. In the real world with billions of people, the answer is less obvious. We take up this issue in the next chapter, where we see that free societies allocate resources through the market forces of supply and demand.

**SUMMARY**

- Each person consumes goods and services produced by many other people both in the United States and around the world. Interdependence and trade are desirable because they allow everyone to enjoy a greater quantity and variety of goods and services.

- There are two ways to compare the ability of two people in producing a good. The person who can produce the good with the smaller quantity of inputs is said to have an absolute advantage in producing the good. The person who has the smaller opportunity cost of producing the good is said to have a comparative advantage. The gains from trade are based on comparative advantage, not absolute advantage.

- Trade makes everyone better off because it allows people to specialize in those activities in which they have a comparative advantage.

- The principle of comparative advantage applies to countries as well as to people. Economists use the principle of comparative advantage to advocate free trade among countries.

**KEY CONCEPTS**

- absolute advantage, p. 54
- comparative advantage, p. 55
- exports, p. 58
- opportunity cost, p. 54
- imports, p. 58
QUESTIONS FOR REVIEW

1. Under what conditions is the production possibilities frontier linear rather than bowed out?
2. Explain how absolute advantage and comparative advantage differ.
3. Give an example in which one person has an absolute advantage in doing something but another person has a comparative advantage.
4. Is absolute advantage or comparative advantage more important for trade? Explain your reasoning using the example in your answer to Question 3.
5. Will a nation tend to export or import goods for which it has a comparative advantage? Explain.
6. Why do economists oppose policies that restrict trade among nations?

PROBLEMS AND APPLICATIONS

1. Maria can read 20 pages of economics in an hour. She can also read 50 pages of sociology in an hour. She spends 5 hours per day studying.
   a. Draw Maria’s production possibilities frontier for reading economics and sociology.
   b. What is Maria’s opportunity cost of reading 100 pages of sociology?
2. American and Japanese workers can each produce 4 cars a year. An American worker can produce 10 tons of grain a year, whereas a Japanese worker can produce 5 tons of grain a year. To keep things simple, assume that each country has 100 million workers.
   a. For this situation, construct a table analogous to the table in Figure 1.
   b. Graph the production possibilities frontier of the American and Japanese economies.
   c. For the United States, what is the opportunity cost of a car? Of grain? For Japan, what is the opportunity cost of a car? Of grain? Put this information in a table analogous to Table 1.
   d. Which country has an absolute advantage in producing cars? In producing grain?
   e. Which country has a comparative advantage in producing cars? In producing grain?
   f. Without trade, half of each country’s workers produce cars and half produce grain. What quantities of cars and grain does each country produce?
   g. Starting from a position without trade, give an example in which trade makes each country better off.
3. Pat and Kris are roommates. They spend most of their time studying (of course), but they leave some time for their favorite activities: making pizza and brewing root beer. Pat takes 4 hours to brew a gallon of root beer and 2 hours to make a pizza. Kris takes 6 hours to brew a gallon of root beer and 4 hours to make a pizza.
   a. What is each roommate’s opportunity cost of making a pizza? Who has the absolute advantage in making pizza? Who has the comparative advantage in making pizza?
   b. If Pat and Kris trade foods with each other, who will trade away pizza in exchange for root beer?
   c. The price of pizza can be expressed in terms of gallons of root beer. What is the highest price at which pizza can be traded that would make both roommates better off? What is the lowest price? Explain.
4. Suppose that there are 10 million workers in Canada and that each of these workers can produce either 2 cars or 30 bushels of wheat in a year.
   a. What is the opportunity cost of producing a car in Canada? What is the opportunity cost of producing a bushel of wheat in Canada? Explain the relationship between the opportunity costs of the two goods.
   b. Draw Canada’s production possibilities frontier. If Canada chooses to consume 10 million cars, how much wheat can it consume without trade? Label this point on the production possibilities frontier.
   c. Now suppose that the United States offers to buy 10 million cars from Canada in exchange for 20 bushels of wheat per car. If Canada continues to consume 10 million cars, how
5. England and Scotland both produce scones and sweaters. Suppose that an English worker can produce 50 scones per hour or 1 sweater per hour. Suppose that a Scottish worker can produce 40 scones per hour or 2 sweaters per hour.
   a. Which country has the absolute advantage in the production of each good? Which country has the comparative advantage?
   b. If England and Scotland decide to trade, which commodity will Scotland trade to England? Explain.
   c. If a Scottish worker could produce only 1 sweater per hour, would Scotland still gain from trade? Would England still gain from trade? Explain.

6. The following table describes the production possibilities of two cities in the country of Baseballia:

<table>
<thead>
<tr>
<th>City</th>
<th>Pairs of Red Socks per Worker per Hour</th>
<th>Pairs of White Socks per Worker per Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Chicago</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

   a. Without trade, what is the price of white socks (in terms of red socks) in Boston? What is the price in Chicago?
   b. Which city has an absolute advantage in the production of each color sock? Which city has a comparative advantage in the production of each color sock?
   c. If the cities trade with each other, which color sock will each export?
   d. What is the range of prices at which trade can occur?

7. Suppose that in a year an American worker can produce 100 shirts or 20 computers, while a Chinese worker can produce 100 shirts or 10 computers.
   a. Graph the production possibilities curve for the two countries. Suppose that without trade the workers in each country spend half their time producing each good. Identify this point in your graph.
   b. If these countries were open to trade, which country would export shirts? Give a specific numerical example and show it on your graph. Which country would benefit from trade? Explain.
   c. Explain at what price of computers (in terms of shirts) the two countries might trade.
   d. Suppose that China catches up with American productivity so that a Chinese worker can produce 100 shirts or 20 computers. What pattern of trade would you predict now? How does this advance in Chinese productivity affect the economic well-being of the citizens of the two countries?

8. An average worker in Brazil can produce an ounce of soybeans in 20 minutes and an ounce of coffee in 60 minutes, while an average worker in Peru can produce an ounce of soybeans in 50 minutes and an ounce of coffee in 75 minutes.
   a. Who has the absolute advantage in coffee? Explain.
   b. Who has the comparative advantage in coffee? Explain.
   c. If the two countries specialize and trade with each other, who will import coffee? Explain.
   d. Assume that the two countries trade and that the country importing coffee trades 2 ounces of soybeans for 1 ounce of coffee. Explain why both countries will benefit from this trade.

9. Are the following statements true or false? Explain in each case.
   a. “Two countries can achieve gains from trade even if one of the countries has an absolute advantage in the production of all goods.”
   b. “Certain very talented people have a comparative advantage in everything they do.”
   c. “If a certain trade is good for one person, it can’t be good for the other one.”
   d. “If a certain trade is good for one person, it is always good for the other one.”
   e. “If trade is good for a country, it must be good for everyone in the country.”

10. The United States exports corn and aircraft to the rest of the world, and it imports oil and clothing from the rest of the world. Do you think this pattern of trade is consistent with the principle of comparative advantage? Why or why not?
PART II
How Markets Work
This page intentionally left blank
The Market Forces of Supply and Demand

When a cold snap hits Florida, the price of orange juice rises in supermarkets throughout the country. When the weather turns warm in New England every summer, the price of hotel rooms in the Caribbean plummets. When a war breaks out in the Middle East, the price of gasoline in the United States rises, and the price of a used Cadillac falls. What do these events have in common? They all show the workings of supply and demand.

Supply and demand are the two words economists use most often—and for good reason. Supply and demand are the forces that make market economies work. They determine the quantity of each good produced and the price at which it is sold. If you want to know how any event or policy will affect the economy, you must think first about how it will affect supply and demand.

This chapter introduces the theory of supply and demand. It considers how buyers and sellers behave and how they interact with one another. It shows how supply and demand determine prices in a market economy and how prices, in turn, allocate the economy’s scarce resources.
MARKETS AND COMPETITION

The terms supply and demand refer to the behavior of people as they interact with one another in competitive markets. Before discussing how buyers and sellers behave, let’s first consider more fully what we mean by the terms market and competition.

**What Is a Market?**

A market is a group of buyers and sellers of a particular good or service. The buyers as a group determine the demand for the product, and the sellers as a group determine the supply of the product.

Markets take many forms. Sometimes markets are highly organized, such as the markets for many agricultural commodities. In these markets, buyers and sellers meet at a specific time and place, where an auctioneer helps set prices and arrange sales.

More often, markets are less organized. For example, consider the market for ice cream in a particular town. Buyers of ice cream do not meet together at any one time. The sellers of ice cream are in different locations and offer somewhat different products. There is no auctioneer calling out the price of ice cream. Each seller posts a price for an ice-cream cone, and each buyer decides how much ice cream to buy at each store. Nonetheless, these consumers and producers of ice cream are closely connected. The ice-cream buyers are choosing from the various ice-cream sellers to satisfy their hunger, and the ice-cream sellers are all trying to appeal to the same ice-cream buyers to make their businesses successful. Even though it is not organized, the group of ice-cream buyers and ice-cream sellers forms a market.

**What Is Competition?**

The market for ice cream, like most markets in the economy, is highly competitive. Each buyer knows that there are several sellers from which to choose, and each seller is aware that his or her product is similar to that offered by other sellers. As a result, the price of ice cream and the quantity of ice cream sold are not determined by any single buyer or seller. Rather, price and quantity are determined by all buyers and sellers as they interact in the marketplace.

Economists use the term competitive market to describe a market in which there are so many buyers and so many sellers that each has a negligible impact on the market price. Each seller of ice cream has limited control over the price because other sellers are offering similar products. A seller has little reason to charge less than the going price, and if he or she charges more, buyers will make their purchases elsewhere. Similarly, no single buyer of ice cream can influence the price of ice cream because each buyer purchases only a small amount.

In this chapter, we assume that markets are perfectly competitive. To reach this highest form of competition, a market must have two characteristics: (1) the goods offered for sale are all exactly the same, and (2) the buyers and sellers are so numerous that no single buyer or seller has any influence over the market price. Because buyers and sellers in perfectly competitive markets must accept the price the market determines, they are said to be price takers. At the market price, buyers can buy all they want, and sellers can sell all they want.
There are some markets in which the assumption of perfect competition applies perfectly. In the wheat market, for example, there are thousands of farmers who sell wheat and millions of consumers who use wheat and wheat products. Because no single buyer or seller can influence the price of wheat, each takes the price as given.

Not all goods and services, however, are sold in perfectly competitive markets. Some markets have only one seller, and this seller sets the price. Such a seller is called a monopoly. Your local cable television company, for instance, may be a monopoly. Residents of your town probably have only one cable company from which to buy this service. Still other markets fall between the extremes of perfect competition and monopoly.

Despite the diversity of market types we find in the world, assuming perfect competition is a useful simplification and, therefore, a natural place to start. Perfectly competitive markets are the easiest to analyze because everyone participating in the market takes the price as given by market conditions. Moreover, because some degree of competition is present in most markets, many of the lessons that we learn by studying supply and demand under perfect competition apply in more complicated markets as well.

**QUICK QUIZ**  What is a market?  •  What are the characteristics of a perfectly competitive market?

---

**DEMAND**

We begin our study of markets by examining the behavior of buyers. To focus our thinking, let’s keep in mind a particular good—ice cream.

**THE DEMAND CURVE: THE RELATIONSHIP BETWEEN PRICE AND QUANTITY DEMANDED**

The quantity demanded of any good is the amount of the good that buyers are willing and able to purchase. As we will see, many things determine the quantity demanded of any good, but when analyzing how markets work, one determinant plays a central role—the price of the good. If the price of ice cream rose to $20 per scoop, you would buy less ice cream. You might buy frozen yogurt instead. If the price of ice cream fell to $0.20 per scoop, you would buy more. This relationship between price and quantity demanded is true for most goods in the economy and, in fact, is so pervasive that economists call it the law of demand: Other things equal, when the price of a good rises, the quantity demanded of the good falls, and when the price falls, the quantity demanded rises.

The table in Figure 1 shows how many ice-cream cones Catherine buys each month at different prices of ice cream. If ice cream is free, Catherine eats 12 cones per month. At $0.50 per cone, Catherine buys 10 cones each month. As the price rises further, she buys fewer and fewer cones. When the price reaches $3.00, Catherine doesn’t buy any ice cream at all. This table is a demand schedule, a table that shows the relationship between the price of a good and the quantity demanded, holding constant everything else that influences how much consumers of the good want to buy.
The graph in Figure 1 uses the numbers from the table to illustrate the law of demand. By convention, the price of ice cream is on the vertical axis, and the quantity of ice cream demanded is on the horizontal axis. The downward-sloping line relating price and quantity demanded is called the demand curve.

**MARKET DEMAND VERSUS INDIVIDUAL DEMAND**

The demand curve in Figure 1 shows an individual’s demand for a product. To analyze how markets work, we need to determine the market demand, the sum of all the individual demands for a particular good or service.

The table in Figure 2 shows the demand schedules for ice cream of the two individuals in this market—Catherine and Nicholas. At any price, Catherine’s demand schedule tells us how much ice cream she buys, and Nicholas’s demand schedule tells us how much ice cream he buys. The market demand at each price is the sum of the two individual demands.

The graph in Figure 2 shows the demand curves that correspond to these demand schedules. Notice that we sum the individual demand curves horizontally to obtain the market demand curve. That is, to find the total quantity demanded at any price, we add the individual quantities, which are found on the horizontal axis of the individual demand curves. Because we are interested in analyzing how markets function, we work most often with the market demand curve. The market demand curve shows how the total quantity demanded of a good varies as the
price of the good varies, while all the other factors that affect how much consumers want to buy are held constant.

**Shifts in the Demand Curve**

Because the market demand curve holds other things constant, it need not be stable over time. If something happens to alter the quantity demanded at any given price, the demand curve shifts. For example, suppose the American Medical Association discovered that people who regularly eat ice cream live longer, healthier lives. The discovery would raise the demand for ice cream. At any given price, buyers would now want to purchase a larger quantity of ice cream, and the demand curve for ice cream would shift.

Figure 3 illustrates shifts in demand. Any change that increases the quantity demanded at every price, such as our imaginary discovery by the American Medical Association, shifts the demand curve to the right and is called an *increase in demand*. Any change that reduces the quantity demanded at every price shifts the demand curve to the left and is called a *decrease in demand*.

There are many variables that can shift the demand curve. Here are the most important.
Income. What would happen to your demand for ice cream if you lost your job one summer? Most likely, it would fall. A lower income means that you have less to spend in total, so you would have to spend less on some—and probably most—goods. If the demand for a good falls when income falls, the good is called a **normal good**.

Not all goods are normal goods. If the demand for a good rises when income falls, the good is called an **inferior good**. An example of an inferior good might be bus rides. As your income falls, you are less likely to buy a car or take a cab and more likely to ride a bus.

**Prices of Related Goods** Suppose that the price of frozen yogurt falls. The law of demand says that you will buy more frozen yogurt. At the same time, you will probably buy less ice cream. Because ice cream and frozen yogurt are both cold, sweet, creamy desserts, they satisfy similar desires. When a fall in the price of one good reduces the demand for another good, the two goods are called **substitutes**. Substitutes are often pairs of goods that are used in place of each other, such as hot dogs and hamburgers, sweaters and sweatshirts, and movie tickets and video rentals.

Now suppose that the price of hot fudge falls. According to the law of demand, you will buy more hot fudge. Yet in this case, you will buy more ice cream as well because ice cream and hot fudge are often used together. When a fall in the price of one good raises the demand for another good, the two goods are called **complements**. Complements are often pairs of goods that are used together, such as gasoline and automobiles, computers and software, and peanut butter and jelly.

**Tastes** The most obvious determinant of your demand is your tastes. If you like ice cream, you buy more of it. Economists normally do not try to explain people’s tastes because tastes are based on historical and psychological forces that are beyond the realm of economics. Economists do, however, examine what happens when tastes change.

---

**normal good**
a good for which, other things equal, an increase in income leads to an increase in demand

**inferior good**
a good for which, other things equal, an increase in income leads to a decrease in demand

**substitutes**
two goods for which an increase in the price of one leads to an increase in the demand for the other

**complements**
two goods for which an increase in the price of one leads to a decrease in the demand for the other
Expectations Your expectations about the future may affect your demand for a good or service today. For example, if you expect to earn a higher income next month, you may choose to save less now and spend more of your current income buying ice cream. As another example, if you expect the price of ice cream to fall tomorrow, you may be less willing to buy an ice-cream cone at today’s price.

Number of Buyers In addition to the preceding factors, which influence the behavior of individual buyers, market demand depends on the number of these buyers. If Peter were to join Catherine and Nicholas as another consumer of ice cream, the quantity demanded in the market would be higher at every price, and market demand would increase.

Summary The demand curve shows what happens to the quantity demanded of a good when its price varies, holding constant all the other variables that influence buyers. When one of these other variables changes, the demand curve shifts. Table 1 lists the variables that influence how much consumers choose to buy of a good. Table 1 lists the variables that affect how much consumers choose to buy of any good. Notice the special role that the price of the good plays: A change in the good’s price represents a movement along the demand curve, whereas a change in one of the other variables shifts the demand curve.

If you have trouble remembering whether you need to shift or move along the demand curve, it helps to recall a lesson from the appendix to Chapter 2. A curve shifts when there is a change in a relevant variable that is not measured on either axis. Because the price is on the vertical axis, a change in price represents a movement along the demand curve. By contrast, income, the prices of related goods, tastes, expectations, and the number of buyers are not measured on either axis, so a change in one of these variables shifts the demand curve.

**TWO WAYS TO REDUCE THE QUANTITY OF SMOKING DEMANDED**

Public policymakers often want to reduce the amount that people smoke. There are two ways that policy can attempt to achieve this goal.

One way to reduce smoking is to shift the demand curve for cigarettes and other tobacco products. Public service announcements, mandatory health warnings on cigarette packages, and the prohibition of cigarette advertising on television are all policies aimed at reducing the quantity of cigarettes demanded at any given price. If successful, these policies shift the demand curve for cigarettes to the left, as in panel (a) of Figure 4.
Alternatively, policymakers can try to raise the price of cigarettes. If the government taxes the manufacture of cigarettes, for example, cigarette companies pass much of this tax on to consumers in the form of higher prices. A higher price encourages smokers to reduce the numbers of cigarettes they smoke. In this case, the reduced amount of smoking does not represent a shift in the demand curve. Instead, it represents a movement along the same demand curve to a point with a higher price and lower quantity, as in panel (b) of Figure 4.

How much does the amount of smoking respond to changes in the price of cigarettes? Economists have attempted to answer this question by studying what happens when the tax on cigarettes changes. They have found that a 10 percent increase in the price causes a 4 percent reduction in the quantity demanded. Teenagers are found to be especially sensitive to the price of cigarettes: A 10 percent increase in the price causes a 12 percent drop in teenage smoking.

A related question is how the price of cigarettes affects the demand for illicit drugs, such as marijuana. Opponents of cigarette taxes often argue that tobacco and marijuana are substitutes so that high cigarette prices encourage marijuana use. By contrast, many experts on substance abuse view tobacco as a “gateway drug” leading the young to experiment with other harmful substances. Most studies of the data are consistent with this latter view: They find that lower cigarette prices are associated with greater use of marijuana. In other words, tobacco and marijuana appear to be complements rather than substitutes.
**QUICK QUIZ**  Make up an example of a monthly demand schedule for pizza and graph the implied demand curve.  
• Give an example of something that would shift this demand curve, and briefly explain your reasoning.  
• Would a change in the price of pizza shift this demand curve?

## SUPPLY

We now turn to the other side of the market and examine the behavior of sellers. Once again, to focus our thinking, let’s consider the market for ice cream.

### THE SUPPLY CURVE: THE RELATIONSHIP BETWEEN PRICE AND QUANTITY SUPPLIED

The *quantity supplied* of any good or service is the amount that sellers are willing and able to sell. There are many determinants of quantity supplied, but once again, price plays a special role in our analysis. When the price of ice cream is high, selling ice cream is profitable, and so the quantity supplied is large. Sellers of ice cream work long hours, buy many ice-cream machines, and hire many workers. By contrast, when the price of ice cream is low, the business is less profitable, and so sellers produce less ice cream. At a low price, some sellers may even choose to shut down, and their quantity supplied falls to zero. This relationship between price and quantity supplied is called the *law of supply*: Other things equal, when the price of a good rises, the quantity supplied of the good also rises, and when the price falls, the quantity supplied falls as well.

The table in Figure 5 shows the quantity of ice-cream cones supplied each month by Ben, an ice-cream seller, at various prices of ice cream. At a price below $1.00, Ben does not supply any ice cream at all. As the price rises, he supplies a greater and greater quantity. This is the *supply schedule*, a table that shows the relationship between the price of a good and the quantity supplied, holding constant everything else that influences how much producers of the good want to sell.

The graph in Figure 5 uses the numbers from the table to illustrate the law of supply. The curve relating price and quantity supplied is called the *supply curve*. The supply curve slopes upward because, other things equal, a higher price means a greater quantity supplied.

### MARKET SUPPLY VERSUS INDIVIDUAL SUPPLY

Just as market demand is the sum of the demands of all buyers, market supply is the sum of the supplies of all sellers. The table in Figure 6 shows the supply schedules for the two ice-cream producers in the market—Ben and Jerry. At any price, Ben’s supply schedule tells us the quantity of ice cream Ben supplies, and Jerry’s supply schedule tells us the quantity of ice cream Jerry supplies. The market supply is the sum of the two individual supplies.

The graph in Figure 6 shows the supply curves that correspond to the supply schedules. As with demand curves, we sum the individual supply curves *horizontally* to obtain the market supply curve. That is, to find the total quantity supplied at any price, we add the individual quantities, which are found on the horizontal axis of the individual supply curves. The market supply curve shows how the total quantity supplied varies as the price of the good varies, holding constant...
all the other factors beyond price that influence producers’ decisions about how much to sell.

**Shifts in the Supply Curve**

Because the market supply curve holds other things constant, the curve shifts when one of the factors changes. For example, suppose the price of sugar falls. Sugar is an input into producing ice cream, so the fall in the price of sugar makes selling ice cream more profitable. This raises the supply of ice cream: At any given price, sellers are now willing to produce a larger quantity. The supply curve for ice cream shifts to the right.

Figure 7 illustrates shifts in supply. Any change that raises quantity supplied at every price, such as a fall in the price of sugar, shifts the supply curve to the right and is called an *increase in supply*. Similarly, any change that reduces the quantity supplied at every price shifts the supply curve to the left and is called a *decrease in supply*.

There are many variables that can shift the supply curve. Here are some of the most important.

**Input Prices** To produce their output of ice cream, sellers use various inputs: cream, sugar, flavoring, ice-cream machines, the buildings in which the ice cream is made, and the labor of workers to mix the ingredients and operate the machines.
Market Supply as the Sum of Individual Supplies

The quantity supplied in a market is the sum of the quantities supplied by all the sellers at each price. Thus, the market supply curve is found by adding horizontally the individual supply curves. At a price of $2.00, Ben supplies 3 ice-cream cones, and Jerry supplies 4 ice-cream cones. The quantity supplied in the market at this price is 7 cones.

Shifts in the Supply Curve

Any change that raises the quantity that sellers wish to produce at any given price shifts the supply curve to the right. Any change that lowers the quantity that sellers wish to produce at any given price shifts the supply curve to the left.
When the price of one or more of these inputs rises, producing ice cream is less profitable, and firms supply less ice cream. If input prices rise substantially, a firm might shut down and supply no ice cream at all. Thus, the supply of a good is negatively related to the price of the inputs used to make the good.

**Technology** The technology for turning inputs into ice cream is another determinant of supply. The invention of the mechanized ice-cream machine, for example, reduced the amount of labor necessary to make ice cream. By reducing firms’ costs, the advance in technology raised the supply of ice cream.

**Expectations** The amount of ice cream a firm supplies today may depend on its expectations about the future. For example, if a firm expects the price of ice cream to rise in the future, it will put some of its current production into storage and supply less to the market today.

**Number of Sellers** In addition to the preceding factors, which influence the behavior of individual sellers, market supply depends on the number of these sellers. If Ben or Jerry were to retire from the ice-cream business, the supply in the market would fall.

**Summary** The supply curve shows what happens to the quantity supplied of a good when its price varies, holding constant all the other variables that influence sellers. When one of these other variables changes, the supply curve shifts. Table 2 lists the variables that influence how much producers choose to sell of a good.

Once again, to remember whether you need to shift or move along the supply curve, keep in mind that a curve shifts only when there is a change in a relevant variable that is not named on either axis. The price is on the vertical axis, so a change in price represents a movement along the supply curve. By contrast, because input prices, technology, expectations, and the number of sellers are not measured on either axis, a change in one of these variables shifts the supply curve.

**Quick Quiz** Make up an example of a monthly supply schedule for pizza and graph the implied supply curve. Give an example of something that would shift this supply curve, and briefly explain your reasoning. Would a change in the price of pizza shift this supply curve?

<table>
<thead>
<tr>
<th>Variable</th>
<th>A Change in This Variable . . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price of the good itself</td>
<td>Represents a movement along the supply curve</td>
</tr>
<tr>
<td>Input prices</td>
<td>Shifts the supply curve</td>
</tr>
<tr>
<td>Technology</td>
<td>Shifts the supply curve</td>
</tr>
<tr>
<td>Expectations</td>
<td>Shifts the supply curve</td>
</tr>
<tr>
<td>Number of sellers</td>
<td>Shifts the supply curve</td>
</tr>
</tbody>
</table>
SUPPLY AND DEMAND TOGETHER

Having analyzed supply and demand separately, we now combine them to see how they determine the price and quantity of a good sold in a market.

EQUILIBRIUM

Figure 8 shows the market supply curve and market demand curve together. Notice that there is one point at which the supply and demand curves intersect. This point is called the market’s equilibrium. The price at this intersection is called the equilibrium price, and the quantity is called the equilibrium quantity. Here the equilibrium price is $2.00 per cone, and the equilibrium quantity is 7 ice-cream cones.

The dictionary defines the word equilibrium as a situation in which various forces are in balance—and this also describes a market’s equilibrium. At the equilibrium price, the quantity of the good that buyers are willing and able to buy exactly balances the quantity that sellers are willing and able to sell. The equilibrium price is sometimes called the market-clearing price because, at this price, everyone in the market has been satisfied: Buyers have bought all they want to buy, and sellers have sold all they want to sell.

The actions of buyers and sellers naturally move markets toward the equilibrium of supply and demand. To see why, consider what happens when the market price is not equal to the equilibrium price.

Suppose first that the market price is above the equilibrium price, as in panel (a) of Figure 9. At a price of $2.50 per cone, the quantity of the good supplied (10 cones) exceeds the quantity demanded (4 cones). There is a surplus of the good: Suppliers are unable to sell all they want at the going price. A surplus is sometimes called a situation of excess supply. When there is a surplus in the ice-cream market,
sellers of ice cream find their freezers increasingly full of ice cream they would like to sell but cannot. They respond to the surplus by cutting their prices. Falling prices, in turn, increase the quantity demanded and decrease the quantity supplied. Prices continue to fall until the market reaches the equilibrium.

Suppose now that the market price is below the equilibrium price, as in panel (b) of Figure 9. In this case, the price is $1.50 per cone, and the quantity of the good demanded (10 cones) exceeds the quantity supplied (4 cones). With too many buyers chasing too few goods, suppliers can take advantage of the shortage by raising the price. Hence, in both cases, the price adjustment moves the market toward the equilibrium of supply and demand.

Thus, the activities of the many buyers and sellers automatically push the market price toward the equilibrium price. Once the market reaches its equilibrium, all buyers and sellers are satisfied, and there is no upward or downward pressure on the price. How quickly equilibrium is reached varies from market to market depending on how quickly prices adjust. In most free markets, surpluses and shortages are only temporary because prices eventually move toward their equilibrium levels. Indeed, this phenomenon is so pervasive that it is called the law of supply and demand: The price of any good adjusts to bring the quantity supplied and quantity demanded for that good into balance.
### Three Steps to Analyzing Changes in Equilibrium

So far, we have seen how supply and demand together determine a market’s equilibrium, which in turn determines the price and quantity of the good that buyers purchase and sellers produce. The equilibrium price and quantity depend on the position of the supply and demand curves. When some event shifts one of these curves, the equilibrium in the market changes, resulting in a new price and a new quantity exchanged between buyers and sellers.

When analyzing how some event affects the equilibrium in a market, we proceed in three steps. First, we decide whether the event shifts the supply curve, the demand curve, or, in some cases, both curves. Second, we decide whether the curve shifts to the right or to the left. Third, we use the supply-and-demand diagram to compare the initial and the new equilibrium, which shows how the shift affects the equilibrium price and quantity. Table 3 summarizes these three steps. To see how this recipe is used, let’s consider various events that might affect the market for ice cream.

**Example: A Change in Market Equilibrium Due to a Shift in Demand**

Suppose that one summer the weather is very hot. How does this event affect the market for ice cream? To answer this question, let’s follow our three steps.

1. The hot weather affects the demand curve by changing people’s taste for ice cream. That is, the weather changes the amount of ice cream that people want to buy at any given price. The supply curve is unchanged because the weather does not directly affect the firms that sell ice cream.

2. Because hot weather makes people want to eat more ice cream, the demand curve shifts to the right. Figure 10 shows this increase in demand as the shift in the demand curve from $D_1$ to $D_2$. This shift indicates that the quantity of ice cream demanded is higher at every price.

3. As Figure 10 shows, the increase in demand raises the equilibrium price from $2.00 to $2.50 and the equilibrium quantity from 7 to 10 cones. In other words, the hot weather increases the price of ice cream and the quantity of ice cream sold.

**Shifts in Curves versus Movements along Curves**

Notice that when hot weather increases the demand for ice cream and drives up the price, the quantity of ice cream that firms supply rises, even though the supply curve remains the same. In this case, economists say there has been an increase in “quantity supplied” but no change in “supply.”

---

### Table 3

**Three Steps for Analyzing Changes in Equilibrium**

1. Decide whether the event shifts the supply or demand curve (or perhaps both).
2. Decide in which direction the curve shifts.
3. Use the supply-and-demand diagram to see how the shift changes the equilibrium price and quantity.
Supply refers to the position of the supply curve, whereas the quantity supplied refers to the amount suppliers wish to sell. In this example, supply does not change because the weather does not alter firms’ desire to sell at any given price. Instead, the hot weather alters consumers’ desire to buy at any given price and thereby shifts the demand curve to the right. The increase in demand causes the equilibrium price to rise. When the price rises, the quantity supplied rises. This increase in quantity supplied is represented by the movement along the supply curve.

To summarize, a shift in the supply curve is called a “change in supply,” and a shift in the demand curve is called a “change in demand.” A movement along a fixed supply curve is called a “change in the quantity supplied,” and a movement along a fixed demand curve is called a “change in the quantity demanded.”

Example: A Change in Market Equilibrium Due to a Shift in Supply  Suppose that during another summer, a hurricane destroys part of the sugarcane crop and drives up the price of sugar. How does this event affect the market for ice cream? Once again, to answer this question, we follow our three steps.

1. The change in the price of sugar, an input into making ice cream, affects the supply curve. By raising the costs of production, it reduces the amount of ice cream that firms produce and sell at any given price. The demand curve does not change because the higher cost of inputs does not directly affect the amount of ice cream households wish to buy.

2. The supply curve shifts to the left because, at every price, the total amount that firms are willing and able to sell is reduced. Figure 11 illustrates this decrease in supply as a shift in the supply curve from $S_1$ to $S_2$. 
3. As Figure 11 shows, the shift in the supply curve raises the equilibrium price from $2.00 to $2.50 and lowers the equilibrium quantity from 7 to 4 cones. As a result of the sugar price increase, the price of ice cream rises, and the quantity of ice cream sold falls.

**Example: Shifts in Both Supply and Demand** Now suppose that a heat wave and a hurricane occur during the same summer. To analyze this combination of events, we again follow our three steps.

1. We determine that both curves must shift. The hot weather affects the demand curve because it alters the amount of ice cream that households want to buy at any given price. At the same time, when the hurricane drives up sugar prices, it alters the supply curve for ice cream because it changes the amount of ice cream that firms want to sell at any given price.

2. The curves shift in the same directions as they did in our previous analysis: The demand curve shifts to the right, and the supply curve shifts to the left. Figure 12 illustrates these shifts.

3. As Figure 12 shows, two possible outcomes might result depending on the relative size of the demand and supply shifts. In both cases, the equilibrium price rises. In panel (a), where demand increases substantially while supply falls just a little, the equilibrium quantity also rises. By contrast, in panel (b), where supply falls substantially while demand rises just a little, the equilibrium quantity falls. Thus, these events certainly raise the price of ice cream, but their impact on the amount of ice cream sold is ambiguous (that is, it could go either way).
Summary  We have just seen three examples of how to use supply and demand curves to analyze a change in equilibrium. Whenever an event shifts the supply curve, the demand curve, or perhaps both curves, you can use these tools to predict how the event will alter the amount sold in equilibrium and the price at which the good is sold. Table 4 shows the predicted outcome for any combination of shifts in the two curves. To make sure you understand how to use the tools of supply and demand, pick a few entries in this table and make sure you can explain to yourself why the table contains the prediction it does.

QUICK QUIZ  On the appropriate diagram, show what happens to the market for pizza if the price of tomatoes rises. • On a separate diagram, show what happens to the market for pizza if the price of hamburgers falls.

<table>
<thead>
<tr>
<th>Type of Graph</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>The pie chart in panel (a) shows how U.S. national income is derived from various sources. The bar graph in panel (b) compares the average income in four countries. The time-series graph in panel (c) shows the productivity of labor in U.S. businesses from 1950 to 2000.</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 4**

<table>
<thead>
<tr>
<th>What Happens to Price and Quantity When Supply or Demand Shifts?</th>
</tr>
</thead>
<tbody>
<tr>
<td>As a quick quiz, make sure you can explain at least a few of the entries in this table using a supply-and-demand diagram.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No Change in Supply</th>
<th>An Increase in Supply</th>
<th>A Decrease in Supply</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Change in Demand</td>
<td>$P_{\text{same}}$</td>
<td>$P_{\text{down}}$</td>
</tr>
<tr>
<td>An Increase in Demand</td>
<td>$P_{\text{up}}$</td>
<td>$P_{\text{ambiguous}}$</td>
</tr>
<tr>
<td>A Decrease in Demand</td>
<td>$P_{\text{down}}$</td>
<td>$P_{\text{ambiguous}}$</td>
</tr>
</tbody>
</table>
As Demand Balloons, Helium Is in Short Supply

By Ana Campoy

Syracuse University physicist Gianfranco Vidali spends most of his time studying how molecules are made in outer space, but a couple of months ago he abruptly dropped his interstellar research to address an earthly issue: the global shortage of helium.

The airy element best known for floating party balloons and the Goodyear blimp is also the lifeblood of a widening world of scientific research. Mr. Vidali uses the gas, which becomes the coldest liquid on earth when pressurized, to recreate conditions similar to outer space. Without it, he can’t work. So when his helium supplier informed him it was cutting deliveries to his lab, Mr. Vidali said, “it sent us into a panic mode.”

Helium is found in varying concentrations in the world’s natural-gas deposits, and is separated out in a special refining process. As with oil and natural gas, the easiest-to-get helium supplies have been tapped and are declining. [A leftward shift in the supply curve.] Meanwhile, scientific research has rapidly multiplied the uses of helium in the past 50 years. [A rightward shift in the demand curve.] It is needed to make computer microchips, flat-panel displays, fiber optics and to operate magnetic resonance imaging, or MRI, scans and welding machines. . . .

Glitches at some of the world’s biggest helium-producing plants have put a further pinch on supplies in the past year. [Another leftward shift in the supply curve.] As supplies have tightened, prices have surged in recent months. For one New York laboratory, prices have increased to $8 a liquid liter, from close to $4 at the end of the summer. [An increase in the equilibrium price.]

The upshot: Helium users—from party planners to welding shops—are having to do with less. [A movement along the demand curve.] Large industrial manufacturers are better able to weather the helium shortage, taking steps like installing equipment that can recycle the gas. So it is the nation’s cash-strapped scientific community that is getting the worst of the crunch.

Soaring helium expenses could shut the doors of some independent labs, many of which have produced important research over the years, and slow down work at bigger research centers. Helium is used in research to find cures to deadly diseases, create new sources of energy and answer questions about how the universe was formed. . . .

Experts predict this situation will eventually price out many helium users, who will find substitutes or modify their technology. Some party balloon businesses are filling balloons with mixtures that contain less helium. Some welders are using argon. Industrial users are installing recovery systems. . . .

Reem Jaafar, a researcher at CUNY, says she will go into another area of physics if helium prices stay at their current levels. “If you have a fixed amount in a grant, and you have to spend it all on helium, you don’t have anything left over,” she says.

CONCLUSION: HOW PRICES ALLOCATE RESOURCES

This chapter has analyzed supply and demand in a single market. Although our discussion has centered on the market for ice cream, the lessons learned here apply in most other markets as well. Whenever you go to a store to buy something,
Price Increases after Natural Disasters

When a natural disaster such as a hurricane hits a region, basic commodities such as gasoline and bottled water experience increasing demand and shrinking supply. These shifts in demand and supply curves cause prices to rise, leading some people to complain about “price gouging.” But, as journalist John Stossel argues in this opinion piece, there is an upside to higher prices after a disaster strikes.

In Praise of Price Gouging
By John Stossel

Politicians and the media are furious about price increases in the wake of Hurricane Katrina. They want gas stations and water sellers punished.

If you want to score points cracking down on mean, greedy profiteers, pushing anti-“gouging” rules is a very good thing. But if you’re one of the people the law “protects” from “price gouging,” you won’t fare as well.

Consider this scenario: You are thirsty—worried that your baby is going to become dehydrated. You find a store that’s open, and the storeowner thinks it’s immoral to take advantage of your distress, so he won’t charge you a dime more than he charged last week. But you can’t buy water from him. It’s sold out.

You continue on your quest, and finally find that dreaded monster, the price gouger. He offers a bottle of water that cost $1 last week at an “outrageous” price—say $20. You pay it to survive the disaster.

You resent the price gouger. But if he hadn’t demanded $20, he’d have been out of water. It was the price gouger’s “exploitation” that saved your child.

It saved her because people look out for their own interests. Before you got to the water seller, other people did. At $1 a bottle, they stocked up. At $20 a bottle, they bought more cautiously. By charging $20, the price gouger makes sure his water goes to those who really need it.

The people the softheaded politicians think are cruelest are doing the most to help. Assuming the demand for bottled water was going to go up, they bought a lot of it, planning to resell it at a steep profit. If they hadn’t done that, that water would not have been available for the people who need it the most.

Might the water have been provided by volunteers? Certainly some people help others out of benevolence. But we can’t count on benevolence. As Adam Smith wrote, “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.”

Consider the store owner’s perspective: If he’s not going to make a big profit, why open up the store at all? Staying in a disaster area is dangerous and means giving up the opportunity to be with family in order to take care of the needs of strangers. Why take the risk?

Any number of services—roofing, for example, carpentry, or tree removal—are in overwhelming demand after a disaster. When the time comes to rebuild New Orleans, it’s safe to predict a shortage of local carpenters: The city’s own population of carpenters won’t be enough.

If this were a totalitarian country, the government might just order a bunch of tradesmen to go to New Orleans. But in a free society, those tradesmen must be persuaded to leave their homes and families, leave their employers and customers, and drive from say, Wisconsin, to take work in New Orleans. If they can’t make more money in Louisiana than Wisconsin, why would they make the trip?

Some may be motivated by a desire to be heroic, but we can’t expect enough heroes to fill the need, week after week; most will travel there for the same reason most Americans go to work: to make money. Any tradesman who treks to a disaster area must get higher pay than he would get in his hometown, or he won’t do the trek. Limit him to what his New Orleans colleagues charged before the storm, and even a would-be hero may say, “the heck with it.”

If he charges enough to justify his venture, he’s likely to be condemned morally or legally by the very people he’s trying to help. But they just don’t understand basic economics. Force prices down, and you keep suppliers out. Let the market work, and competition brings prices as low as the challenges of the disaster allow. Goods that were in short supply become available, even to the poor.

It’s the price “gougers” who bring the water, ship the gasoline, fix the roof, and rebuild the cities. The price “gougers” save lives.

you are contributing to the demand for that item. Whenever you look for a job, you are contributing to the supply of labor services. Because supply and demand are such pervasive economic phenomena, the model of supply and demand is a powerful tool for analysis. We will be using this model repeatedly in the following chapters.

One of the Ten Principles of Economics discussed in Chapter 1 is that markets are usually a good way to organize economic activity. Although it is still too early to judge whether market outcomes are good or bad, in this chapter we have begun to see how markets work. In any economic system, scarce resources have to be allocated among competing uses. Market economies harness the forces of supply and demand to serve that end. Supply and demand together determine the prices of the economy’s many different goods and services; prices in turn are the signals that guide the allocation of resources.

For example, consider the allocation of beachfront land. Because the amount of this land is limited, not everyone can enjoy the luxury of living by the beach. Who gets this resource? The answer is whoever is willing and able to pay the price. The price of beachfront land adjusts until the quantity of land demanded exactly balances the quantity supplied. Thus, in market economies, prices are the mechanism for rationing scarce resources.

Similarly, prices determine who produces each good and how much is produced. For instance, consider farming. Because we need food to survive, it is crucial that some people work on farms. What determines who is a farmer and who is not? In a free society, there is no government planning agency making this decision and ensuring an adequate supply of food. Instead, the allocation of workers to farms is based on the job decisions of millions of workers. This decentralized system works well because these decisions depend on prices. The prices of food and the wages of farmworkers (the price of their labor) adjust to ensure that enough people choose to be farmers.

If a person had never seen a market economy in action, the whole idea might seem preposterous. Economies are enormous groups of people engaged in a multitude of interdependent activities. What prevents decentralized decision making from degenerating into chaos? What coordinates the actions of the millions of people with their varying abilities and desires? What ensures that what needs to be done is in fact done? The answer, in a word, is prices. If an invisible hand guides market economies, as Adam Smith famously suggested, then the price system is the baton that the invisible hand uses to conduct the economic orchestra.

**SUMMARY**

- Economists use the model of supply and demand to analyze competitive markets. In a competitive market, there are many buyers and sellers, each of whom has little or no influence on the market price.
- The demand curve shows how the quantity of a good demanded depends on the price. According to the law of demand, as the price of a good falls, the quantity demanded rises. Therefore, the demand curve slopes downward.
- In addition to price, other determinants of how much consumers want to buy include income, the prices of substitutes and complements, tastes, expectations, and the number of buyers. If one of these factors changes, the demand curve shifts.
• The supply curve shows how the quantity of a good supplied depends on the price. According to the law of supply, as the price of a good rises, the quantity supplied rises. Therefore, the supply curve slopes upward.

• In addition to price, other determinants of how much producers want to sell include input prices, technology, expectations, and the number of sellers. If one of these factors changes, the supply curve shifts.

• The intersection of the supply and demand curves determines the market equilibrium. At the equilibrium price, the quantity demanded equals the quantity supplied.

• The behavior of buyers and sellers naturally drives markets toward their equilibrium. When the market price is above the equilibrium price, there is a surplus of the good, which causes the market price to fall. When the market price is below the equilibrium price, there is a shortage, which causes the market price to rise.

• To analyze how any event influences a market, we use the supply-and-demand diagram to examine how the event affects the equilibrium price and quantity. To do this, we follow three steps. First, we decide whether the event shifts the supply curve or the demand curve (or both). Second, we decide in which direction the curve shifts. Third, we compare the new equilibrium with the initial equilibrium.

• In market economies, prices are the signals that guide economic decisions and thereby allocate scarce resources. For every good in the economy, the price ensures that supply and demand are in balance. The equilibrium price then determines how much of the good buyers choose to consume and how much sellers choose to produce.

KEY CONCEPTS

market, p. 66
competitive market, p. 66
quantity demanded, p. 67
law of demand, p. 67
demand schedule, p. 67
demand curve, p. 67
normal good, p. 70
inferior good, p. 70
substitutes, p. 70
complements, p. 70
quantity supplied, p. 73
law of supply, p. 73
supply schedule, p. 73
supply curve, p. 73
equilibrium, p. 77
equilibrium price, p. 77
equilibrium quantity, p. 77
surplus, p. 77
shortage, p. 78
law of supply and demand, p. 78

QUESTIONS FOR REVIEW

1. What is a competitive market? Briefly describe a type of market that is not perfectly competitive.
2. What are the demand schedule and the demand curve and how are they related? Why does the demand curve slope downward?
3. Does a change in consumers’ tastes lead to a movement along the demand curve or a shift in the demand curve? Does a change in price lead to a movement along the demand curve or a shift in the demand curve?
4. Popeye’s income declines, and as a result, he buys more spinach. Is spinach an inferior or a normal good? What happens to Popeye’s demand curve for spinach?
5. What are the supply schedule and the supply curve and how are they related? Why does the supply curve slope upward?
6. Does a change in producers’ technology lead to a movement along the supply curve or a shift in the supply curve? Does a change in price lead to a movement along the supply curve or a shift in the supply curve?
7. Define the equilibrium of a market. Describe the forces that move a market toward its equilibrium.
8. Beer and pizza are complements because they are often enjoyed together. When the price of beer rises, what happens to the supply, demand, quantity supplied, quantity demanded, and the price in the market for pizza?

9. Describe the role of prices in market economies.

**PROBLEMS AND APPLICATIONS**

1. Explain each of the following statements using supply-and-demand diagrams.
   a. “When a cold snap hits Florida, the price of orange juice rises in supermarkets throughout the country.”
   b. “When the weather turns warm in New England every summer, the price of hotel rooms in Caribbean resorts plummeets.”
   c. “When a war breaks out in the Middle East, the price of gasoline rises, and the price of a used Cadillac falls.”
2. “An increase in the demand for notebooks raises the quantity of notebooks demanded but not the quantity supplied.” Is this statement true or false? Explain.
3. Consider the market for minivans. For each of the events listed here, identify which of the determinants of demand or supply are affected. Also indicate whether demand or supply increases or decreases. Then draw a diagram to show the effect on the price and quantity of minivans.
   a. People decide to have more children.
   b. A strike by steelworkers raises steel prices.
   c. Engineers develop new automated machinery for the production of minivans.
   d. The price of sports utility vehicles rises.
   e. A stock-market crash lowers people’s wealth.
4. Identify the flaw in this analysis: “If more Americans go on a low-carb diet, the demand for bread will fall. The decrease in the demand for bread will cause the price of bread to fall. The lower price, however, will then increase the demand. In the new equilibrium, Americans might end up consuming more bread than they did initially.”
5. Consider the markets for DVD movies, TV screens, and tickets at movie theaters.
   a. For each pair, identify whether they are complements or substitutes:
       - DVDs and TV screens
       - DVDs and movie tickets
       - TV screens and movie tickets
   b. Suppose a technological advance reduces the cost of manufacturing TV screens. Draw a diagram to show what happens in the market for TV screens.
   c. Draw two more diagrams to show how the change in the market for TV screens affects the markets for DVDs and movie tickets.
6. Over the past 20 years, technological advances have reduced the cost of computer chips. How do you think this affected the market for computers? For computer software? For typewriters?
7. Using supply-and-demand diagrams, show the effect of the following events on the market for sweatshirts.
   a. A hurricane in South Carolina damages the cotton crop.
   b. The price of leather jackets falls.
   c. All colleges require morning exercise in appropriate attire.
   d. New knitting machines are invented.
8. A survey shows an increase in drug use by young people. In the ensuing debate, two hypotheses are proposed:
   - Reduced police efforts have increased the availability of drugs on the street.
   - Cutbacks in education efforts have decreased awareness of the dangers of drug addiction.
   a. Use supply-and-demand diagrams to show how each of these hypotheses could lead to an increase in quantity of drugs consumed.
   b. How could information on what has happened to the price of drugs help us to distinguish between these explanations?
9. Suppose that in the year 2010 the number of births is temporarily high. How does this baby boom affect the price of babysitting services in 2015 and 2025? (Hint: 5-year-olds need babysitters, whereas 15-year-olds can be babysitters.)
10. Ketchup is a complement (as well as a condiment) for hot dogs. If the price of hot dogs rises, what happens to the market for ketchup? For tomatoes? For tomato juice? For orange juice?
Currently, the demand and supply schedules are as follows:

<table>
<thead>
<tr>
<th>Price</th>
<th>Quantity Demanded</th>
<th>Quantity Supplied</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4</td>
<td>10,000 tickets</td>
<td>8,000 tickets</td>
</tr>
<tr>
<td>8</td>
<td>8,000</td>
<td>8,000</td>
</tr>
<tr>
<td>12</td>
<td>6,000</td>
<td>8,000</td>
</tr>
<tr>
<td>16</td>
<td>4,000</td>
<td>8,000</td>
</tr>
<tr>
<td>20</td>
<td>2,000</td>
<td>8,000</td>
</tr>
</tbody>
</table>

a. Draw the demand and supply curves. What is unusual about this supply curve? Why might this be true?
b. What are the equilibrium price and quantity of tickets?
c. Your college plans to increase total enrollment next year by 5,000 students. The additional students will have the following demand schedule:

<table>
<thead>
<tr>
<th>Price</th>
<th>Quantity Demanded</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4</td>
<td>4,000 tickets</td>
</tr>
<tr>
<td>8</td>
<td>3,000</td>
</tr>
<tr>
<td>12</td>
<td>2,000</td>
</tr>
<tr>
<td>16</td>
<td>1,000</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
</tr>
</tbody>
</table>

Now add the old demand schedule and the demand schedule for the new students to calculate the new demand schedule for the entire college. What will be the new equilibrium price and quantity?

15. Market research has revealed the following information about the market for chocolate bars: The demand schedule can be represented by the equation \( Q^D = 1,600 - 300P \), where \( Q^D \) is the quantity demanded and \( P \) is the price. The supply schedule can be represented by the equation \( Q^S = 1,400 + 700P \), where \( Q^S \) is the quantity supplied. Calculate the equilibrium price and quantity in the market for chocolate bars.
Imagine that some event drives up the price of gasoline in the United States. It could be a war in the Middle East that disrupts the world supply of oil, a booming Chinese economy that boosts the world demand for oil, or a new tax on gasoline passed by Congress. How would U.S. consumers respond to the higher price?

It is easy to answer this question in broad fashion: Consumers would buy less. That is simply the law of demand we learned in the previous chapter. But you might want a precise answer. By how much would consumption of gasoline fall? This question can be answered using a concept called elasticity, which we develop in this chapter.

Elasticity is a measure of how much buyers and sellers respond to changes in market conditions. When studying how some event or policy affects a market, we can discuss not only the direction of the effects but their magnitude as well. Elasticity is useful in many applications, as we will see toward the end of this chapter.

Before proceeding, however, you might be curious about the answer to the gasoline question. Many studies have examined consumers’ response to gasoline prices, and they typically find that the quantity demanded responds more in the
THE ELASTICITY OF DEMAND

When we introduced demand in Chapter 4, we noted that consumers usually buy more of a good when its price is lower, when their incomes are higher, when the prices of substitutes for the good are higher, or when the prices of complements of the good are lower. Our discussion of demand was qualitative, not quantitative. That is, we discussed the direction in which quantity demanded moves but not the size of the change. To measure how much consumers respond to changes in these variables, economists use the concept of elasticity.

THE PRICE ELASTICITY OF DEMAND AND ITS DETERMINANTS

The law of demand states that a fall in the price of a good raises the quantity demanded. The price elasticity of demand measures how much the quantity demanded responds to a change in price. Demand for a good is said to be elastic if the quantity demanded responds substantially to changes in the price. Demand is said to be inelastic if the quantity demanded responds only slightly to changes in the price.

The price elasticity of demand for any good measures how willing consumers are to buy less of the good as its price rises. Thus, the elasticity reflects the many economic, social, and psychological forces that shape consumer preferences. Based on experience, however, we can state some general rules about what determines the price elasticity of demand.

Availability of Close Substitutes Goods with close substitutes tend to have more elastic demand because it is easier for consumers to switch from that good to others. For example, butter and margarine are easily substitutable. A small increase in the price of butter, assuming the price of margarine is held fixed, causes the quantity of butter sold to fall by a large amount. By contrast, because eggs are a food without a close substitute, the demand for eggs is less elastic than the demand for butter.

Necessities versus Luxuries Necessities tend to have inelastic demands, whereas luxuries have elastic demands. When the price of a doctor’s visit rises, people will not dramatically reduce the number of times they go to the doctor, although they might go somewhat less often. By contrast, when the price of sailboats rises, the quantity of sailboats demanded falls substantially. The reason is that most people view doctor visits as a necessity and sailboats as a luxury. Of course, whether a good is a necessity or a luxury depends not on the intrinsic properties of the good but on the preferences of the buyer. For avid sailors with little concern over their health, sailboats might be a necessity with inelastic demand and doctor visits a luxury with elastic demand.

long run than it does in the short run. A 10 percent increase in gasoline prices reduces gasoline consumption by about 2.5 percent after a year and about 6 percent after five years. About half of the long-run reduction in quantity demanded arises because people drive less and half because they switch to more fuel-efficient cars. Both responses are reflected in the demand curve and its elasticity.
Definition of the Market  The elasticity of demand in any market depends on how we draw the boundaries of the market. Narrowly defined markets tend to have more elastic demand than broadly defined markets because it is easier to find close substitutes for narrowly defined goods. For example, food, a broad category, has a fairly inelastic demand because there are no good substitutes for food. Ice cream, a narrower category, has a more elastic demand because it is easy to substitute other desserts for ice cream. Vanilla ice cream, a very narrow category, has a very elastic demand because other flavors of ice cream are almost perfect substitutes for vanilla.

Time Horizon  Goods tend to have more elastic demand over longer time horizons. When the price of gasoline rises, the quantity of gasoline demanded falls only slightly in the first few months. Over time, however, people buy more fuel-efficient cars, switch to public transportation, and move closer to where they work. Within several years, the quantity of gasoline demanded falls more substantially.

Computing the Price Elasticity of Demand

Now that we have discussed the price elasticity of demand in general terms, let’s be more precise about how it is measured. Economists compute the price elasticity of demand as the percentage change in the quantity demanded divided by the percentage change in the price. That is,

\[
\text{Price elasticity of demand} = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in price}}.
\]

For example, suppose that a 10 percent increase in the price of an ice-cream cone causes the amount of ice cream you buy to fall by 20 percent. We calculate your elasticity of demand as

\[
\text{Price elasticity of demand} = \frac{20 \text{ percent}}{10 \text{ percent}} = 2.
\]

In this example, the elasticity is 2, reflecting that the change in the quantity demanded is proportionately twice as large as the change in the price.

Because the quantity demanded of a good is negatively related to its price, the percentage change in quantity will always have the opposite sign as the percentage change in price. In this example, the percentage change in price is a positive 10 percent (reflecting an increase), and the percentage change in quantity demanded is a negative 20 percent (reflecting a decrease). For this reason, price elasticities of demand are sometimes reported as negative numbers. In this book, we follow the common practice of dropping the minus sign and reporting all price elasticities of demand as positive numbers. (Mathematicians call this the absolute value.) With this convention, a larger price elasticity implies a greater responsiveness of quantity demanded to price.

The Midpoint Method: A Better Way to Calculate Percentage Changes and Elasticities

If you try calculating the price elasticity of demand between two points on a demand curve, you will quickly notice an annoying problem: The elasticity from
point A to point B seems different from the elasticity from point B to point A. For example, consider these numbers:

<table>
<thead>
<tr>
<th>Point</th>
<th>Price</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point A</td>
<td>$4</td>
<td>120</td>
</tr>
<tr>
<td>Point B</td>
<td>$6</td>
<td>80</td>
</tr>
</tbody>
</table>

Going from point A to point B, the price rises by 50 percent, and the quantity falls by 33 percent, indicating that the price elasticity of demand is 33/50, or 0.66. By contrast, going from point B to point A, the price falls by 33 percent, and the quantity rises by 50 percent, indicating that the price elasticity of demand is 50/33, or 1.5. This difference arises because the percentage changes are calculated from a different base.

One way to avoid this problem is to use the midpoint method for calculating elasticities. The standard procedure for computing a percentage change is to divide the change by the initial level. By contrast, the midpoint method computes a percentage change by dividing the change by the midpoint (or average) of the initial and final levels. For instance, $5 is the midpoint between $4 and $6. Therefore, according to the midpoint method, a change from $4 to $6 is considered a 40 percent rise because \[(6 - 4) / 5 \times 100 = 40\]. Similarly, a change from $6 to $4 is considered a 40 percent fall.

Because the midpoint method gives the same answer regardless of the direction of change, it is often used when calculating the price elasticity of demand between two points. In our example, the midpoint between point A and point B is:

<table>
<thead>
<tr>
<th>Midpoint</th>
<th>Price</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$5</td>
<td>100</td>
</tr>
</tbody>
</table>

According to the midpoint method, when going from point A to point B, the price rises by 40 percent, and the quantity falls by 40 percent. Similarly, when going from point B to point A, the price falls by 40 percent, and the quantity rises by 40 percent. In both directions, the price elasticity of demand equals 1.

The following formula expresses the midpoint method for calculating the price elasticity of demand between two points, denoted \((Q_1, P_1)\) and \((Q_2, P_2)\):

\[
\text{Price elasticity of demand} = \frac{(Q_2 - Q_1)}{(P_2 - P_1)} \times \frac{1}{\frac{(Q_2 + Q_1)}{2} / \frac{(P_2 + P_1)}{2}}.
\]

The numerator is the percentage change in quantity computed using the midpoint method, and the denominator is the percentage change in price computed using the midpoint method. If you ever need to calculate elasticities, you should use this formula.

In this book, however, we rarely perform such calculations. For most of our purposes, what elasticity represents—the responsiveness of quantity demanded to a change in price—is more important than how it is calculated.

**The Variety of Demand Curves**

Economists classify demand curves according to their elasticity. Demand is considered *elastic* when the elasticity is greater than 1, which means the quantity moves proportionately more than the price. Demand is considered *inelastic* when the elasticity is less than 1, which means the quantity moves proportionately less
The price elasticity of demand determines whether the demand curve is steep or flat. Note that all percentage changes are calculated using the midpoint method.

**FIGURE 1**

- **(a) Perfectly Inelastic Demand: Elasticity Equals 0**
  - 1. An increase in price...
  - 2. . . . leaves the quantity demanded unchanged.

- **(b) Inelastic Demand: Elasticity Is Less Than 1**
  - 1. A 22% increase in price...
  - 2. . . . leads to an 11% decrease in quantity demanded.

- **(c) Unit Elastic Demand: Elasticity Equals 1**
  - 1. A 22% increase in price...
  - 2. . . . leads to a 22% decrease in quantity demanded.

- **(d) Elastic Demand: Elasticity Is Greater Than 1**
  - 1. A 22% increase in price...
  - 2. . . . leads to a 67% decrease in quantity demanded.

- **(e) Perfectly Elastic Demand: Elasticity Equals Infinity**
  - 1. At any price above $4, quantity demanded is zero.
  - 2. At exactly $4, consumers will buy any quantity.
  - 3. At a price below $4, quantity demanded is infinite.
than the price. If the elasticity is exactly 1, the quantity moves the same amount proportionately as the price, and demand is said to have unit elasticity.

Because the price elasticity of demand measures how much quantity demanded responds to changes in the price, it is closely related to the slope of the demand curve. The following rule of thumb is a useful guide: The flatter the demand curve that passes through a given point, the greater the price elasticity of demand. The steeper the demand curve that passes through a given point, the smaller the price elasticity of demand.

Figure 1 on the previous page shows five cases. In the extreme case of a zero elasticity, shown in panel (a), demand is perfectly inelastic, and the demand curve is vertical. In this case, regardless of the price, the quantity demanded stays the same. As the elasticity rises, the demand curve gets flatter and flatter, as shown in panels (b), (c), and (d). At the opposite extreme, shown in panel (e), demand is perfectly elastic. This occurs as the price elasticity of demand approaches infinity and the demand curve becomes horizontal, reflecting the fact that very small changes in the price lead to huge changes in the quantity demanded.

Finally, if you have trouble keeping straight the terms elastic and inelastic, here’s a memory trick for you: Inelastic curves, such as in panel (a) of Figure 1, look like the letter I. This is not a deep insight, but it might help on your next exam.

**Total Revenue and the Price Elasticity of Demand**

When studying changes in supply or demand in a market, one variable we often want to study is total revenue, the amount paid by buyers and received by sellers of the good. In any market, total revenue is $P \times Q$, the price of the good times the quantity of the good sold. We can show total revenue graphically, as in Figure 2. The height of the box under the demand curve is $P$, and the width is $Q$. The area of this box, $P \times Q$, equals the total revenue in this market. In Figure 2, where $P = $4 and $Q = 100$, total revenue is $4 \times 100$, or $400$.

How does total revenue change as one moves along the demand curve? The answer depends on the price elasticity of demand. If demand is inelastic, as in panel (a) of Figure 3, then an increase in the price causes an increase in total revenue. Here an increase in price from $1$ to $3$ causes the quantity demanded to fall from 100 to 80, so total revenue rises from $100$ to $240$. An increase in price raises $P \times Q$ because the fall in $Q$ is proportionately smaller than the rise in $P$.

We obtain the opposite result if demand is elastic: An increase in the price causes a decrease in total revenue. In panel (b) of Figure 3, for instance, when the price rises from $4$ to $5$, the quantity demanded falls from 50 to 20, so total revenue falls from $200$ to $100$. Because demand is elastic, the reduction in the quantity demanded is so great that it more than offsets the increase in the price. That is, an increase in price reduces $P \times Q$ because the fall in $Q$ is proportionately greater than the rise in $P$.

Although the examples in this figure are extreme, they illustrate some general rules:

- When demand is inelastic (a price elasticity less than 1), price and total revenue move in the same direction.
- When demand is elastic (a price elasticity greater than 1), price and total revenue move in opposite directions.
• If demand is unit elastic (a price elasticity exactly equal to 1), total revenue remains constant when the price changes.

**Elasticity and Total Revenue along a Linear Demand Curve**

Let’s examine how elasticity varies along a linear demand curve, as shown in Figure 4. We know that a straight line has a constant slope. Slope is defined as “rise over run,” which here is the ratio of the change in price (“rise”) to the change in quantity (“run”). This particular demand curve’s slope is constant because each $1 increase in price causes the same two-unit decrease in the quantity demanded.

Even though the slope of a linear demand curve is constant, the elasticity is not. This is true because the slope is the ratio of changes in the two variables, whereas the elasticity is the ratio of percentage changes in the two variables. You can see this by looking at the table in Figure 4, which shows the demand schedule for the linear demand curve in the graph. The table uses the midpoint method to calculate the price elasticity of demand. At points with a low price and high quantity, the demand curve is inelastic. At points with a high price and low quantity, the demand curve is elastic.

The table also presents total revenue at each point on the demand curve. These numbers illustrate the relationship between total revenue and elasticity. When the price is $1, for instance, demand is inelastic, and a price increase to $2 raises total revenue. When the price is $5, demand is elastic, and a price increase to $6 reduces total revenue. Between $3 and $4, demand is exactly unit elastic, and total revenue is the same at these two prices.

The linear demand curve illustrates that the price elasticity of demand need not be the same at all points on a demand curve. A constant elasticity is possible, but it is not always the case.
The impact of a price change on total revenue (the product of price and quantity) depends on the elasticity of demand. In panel (a), the demand curve is inelastic. In this case, an increase in the price leads to a decrease in quantity demanded that is proportionately smaller, so total revenue increases. Here an increase in the price from $1 to $3 causes the quantity demanded to fall from 100 to 80. Total revenue rises from $100 to $240. In panel (b), the demand curve is elastic. In this case, an increase in the price leads to a decrease in quantity demanded that is proportionately larger, so total revenue decreases. Here an increase in the price from $4 to $5 causes the quantity demanded to fall from 50 to 20. Total revenue falls from $200 to $100.
Other Demand Elasticities

In addition to the price elasticity of demand, economists use other elasticities to describe the behavior of buyers in a market.

The Income Elasticity of Demand The income elasticity of demand measures how the quantity demanded changes as consumer income changes. It is calculated as the percentage change in quantity demanded divided by the percentage change in income. That is,

\[
\text{Income elasticity of demand} = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in income}}.
\]

As we discussed in Chapter 4, most goods are normal goods: Higher income raises the quantity demanded. Because quantity demanded and income move in the same direction, normal goods have positive income elasticities. A few goods, such as bus rides, are inferior goods: Higher income lowers the quantity demanded. Because quantity demanded and income move in opposite directions, inferior goods have negative income elasticities.
Real Energy Savers
Don’t Wear Cardigans. Or Do They?
By Anna Bernasek

When oil and gas prices surged after Hurricanes Katrina and Rita, President Bush appealed to Americans to conserve energy. He asked people to cut back on nonessential travel, for example, and to carpool to work. Then, in October, the White House started a campaign for energy conservation in American homes, dusting off some old ideas like switching to fluorescent light bulbs and installing better insulation in attics.

Some critics derided the program as a bizarre flashback from the 1970’s—a collection of worn-out ideas that evoked feelings of deprivation and gloom. It will be a pity, though, if an effective energy policy never gets off the ground. Much has been learned since the 70’s about what works and what doesn’t.

There are reasons for optimism. One is that market forces can help provide solutions: higher prices, on their own, can make people cut back. Just how responsive consumers are to price changes—what economists call the elasticity of demand—has been the focus of much research. Today, economists believe that they have developed a pretty good rule of thumb for energy use. In the case of electricity, which is relatively easy to measure, they have found that when the price rises 10 percent, electricity use falls roughly 3 percent. At the gas pump, a 10 percent increase in price leads to a decline of around 2 percent in demand. [Author’s note: It would be more precise to say that the price increase leads to a 2 percent decline in quantity demanded, because the change represents a movement along the demand curve.]

Consumer behavior can change quickly in a crisis. A study by Peter C. Reiss, a professor of economics at Stanford, and Matthew W. White, a professor of business and public policy at the Wharton School of the University of Pennsylvania, provides some recent evidence. In examining San Diego households during the California electricity crisis of 2000 and 2001, they found that use of electricity dropped surprisingly fast. In the summer of 2000, within 60 days of seeing monthly electric bills rise by about $60—an increase of 130 percent—the average household cut its use of electricity by 12 percent.

That kind of drop requires a big change in behavior. The authors found that households had turned off air-conditioners in the middle of summer and had invested in new energy-efficient appliances, among other things.

High costs aren’t the only force that will influence consumers to cut back. Although public appeals to save energy may be ridiculed by comedians on late-night television—recall President Jimmy Carter’s cardigan sweater—the efforts can have a substantial impact.

Professors Reiss and White found that to be true in San Diego. In February 2001, with electricity prices capped, the state of California began a campaign to have households conserve electricity. It worked. “It was clear by about six months into 2001 that public appeals were having a big impact,” Professor White said. Such campaigns can have significant effects on consumer behavior, he said, if they offer a clear explanation of what people can do and how it will make a difference.

Perhaps the most important reason for optimism is technology’s role in promoting energy savings. From 1979 to 1985, in the aftermath of energy shortages, Americans reduced their oil consumption by 15 percent. The single biggest factor was a shift in car-buying habits. Americans found that driving fuel-efficient cars, instead of gas guzzlers, didn’t stop them from going where they wanted to go.

Even among normal goods, income elasticities vary substantially in size. Necessities, such as food and clothing, tend to have small income elasticities because consumers choose to buy some of these goods even when their incomes are low. Luxuries, such as caviar and diamonds, tend to have large income elasticities because consumers feel that they can do without these goods altogether if their incomes are too low.

The Cross-Price Elasticity of Demand  The cross-price elasticity of demand measures how the quantity demanded of one good responds to a change in the price of another good. It is calculated as the percentage change in quantity demanded of good 1 divided by the percentage change in the price of good 2. That is,

\[
\text{Cross-price elasticity of demand} = \frac{\text{Percentage change in quantity demanded of good 1}}{\text{Percentage change in the price of good 2}}.
\]

Whether the cross-price elasticity is a positive or negative number depends on whether the two goods are substitutes or complements. As we discussed in Chapter 4, substitutes are goods that are typically used in place of one another, such as hamburgers and hot dogs. An increase in hot dog prices induces people to grill hamburgers instead. Because the price of hot dogs and the quantity of hamburgers demanded move in the same direction, the cross-price elasticity is positive. Conversely, complements are goods that are typically used together, such as computers and software. In this case, the cross-price elasticity is negative, indicating that an increase in the price of computers reduces the quantity of software demanded.

Quick Quiz  Define the price elasticity of demand. • Explain the relationship between total revenue and the price elasticity of demand.

The Elasticity of Supply

When we introduced supply in Chapter 4, we noted that producers of a good offer to sell more of it when the price of the good rises. To turn from qualitative to quantitative statements about quantity supplied, we once again use the concept of elasticity.

The Price Elasticity of Supply and Its Determinants

The law of supply states that higher prices raise the quantity supplied. The price elasticity of supply measures how much the quantity supplied responds to changes in the price. Supply of a good is said to be elastic if the quantity supplied responds substantially to changes in the price. Supply is said to be inelastic if the quantity supplied responds only slightly to changes in the price.

The price elasticity of supply depends on the flexibility of sellers to change the amount of the good they produce. For example, beachfront land has an inelastic supply because it is almost impossible to produce more of it. By contrast, manufactured goods, such as books, cars, and televisions, have elastic supplies because
Firms that produce them can run their factories longer in response to a higher price.

In most markets, a key determinant of the price elasticity of supply is the time period being considered. Supply is usually more elastic in the long run than in the short run. Over short periods of time, firms cannot easily change the size of their factories to make more or less of a good. Thus, in the short run, the quantity supplied is not very responsive to the price. By contrast, over longer periods, firms can build new factories or close old ones. In addition, new firms can enter a market, and old firms can shut down. Thus, in the long run, the quantity supplied can respond substantially to price changes.

**Computing the Price Elasticity of Supply**

Now that we have a general understanding about the price elasticity of supply, let’s be more precise. Economists compute the price elasticity of supply as the percentage change in the quantity supplied divided by the percentage change in the price. That is,

\[
\text{Price elasticity of supply} = \frac{\text{Percentage change in quantity supplied}}{\text{Percentage change in price}}
\]

For example, suppose that an increase in the price of milk from $2.85 to $3.15 a gallon raises the amount that dairy farmers produce from 9,000 to 11,000 gallons per month. Using the midpoint method, we calculate the percentage change in price as

\[
\text{Percentage change in price} = \frac{(3.15 - 2.85)}{3.00} \times 100 = 10\%.
\]

Similarly, we calculate the percentage change in quantity supplied as

\[
\text{Percentage change in quantity supplied} = \frac{(11,000 - 9,000)}{10,000} \times 100 = 20\%.
\]

In this case, the price elasticity of supply is

\[
\text{Price elasticity of supply} = \frac{20\%}{10\%} = 2.0.
\]

In this example, the elasticity of 2 indicates that the quantity supplied changes proportionately twice as much as the price.

**The Variety of Supply Curves**

Because the price elasticity of supply measures the responsiveness of quantity supplied to the price, it is reflected in the appearance of the supply curve. Figure 5 shows five cases. In the extreme case of a zero elasticity, as shown in panel (a), supply is *perfectly inelastic*, and the supply curve is vertical. In this case, the quantity supplied is the same regardless of the price. As the elasticity rises, the supply curve gets flatter, which shows that the quantity supplied responds more to changes in the price. At the opposite extreme, shown in panel (e), supply is *perfectly elastic*. This occurs as the price elasticity of supply approaches infinity and the supply curve becomes horizontal, meaning that very small changes in the price lead to very large changes in the quantity supplied.
The price elasticity of supply determines whether the supply curve is steep or flat. Note that all percentage changes are calculated using the midpoint method.

(a) Perfectly Inelastic Supply: Elasticity Equals 0

1. An increase in price . . .
2. . . . leaves the quantity supplied unchanged.

(b) Inelastic Supply: Elasticity Is Less Than 1

1. A 22% increase in price . . .
2. . . . leads to a 10% increase in quantity supplied.

(c) Unit Elastic Supply: Elasticity Equals 1

1. A 22% increase in price . . .
2. . . . leads to a 22% increase in quantity supplied.

(d) Elastic Supply: Elasticity Is Greater Than 1

1. A 22% increase in price . . .
2. . . . leads to a 67% increase in quantity supplied.

(e) Perfectly Elastic Supply: Elasticity Equals Infinity

1. At any price above $4, quantity supplied is infinite.
2. At exactly $4, producers will supply any quantity.
3. At a price below $4, quantity supplied is zero.
In some markets, the elasticity of supply is not constant but varies over the supply curve. Figure 6 shows a typical case for an industry in which firms have factories with a limited capacity for production. For low levels of quantity supplied, the elasticity of supply is high, indicating that firms respond substantially to changes in the price. In this region, firms have capacity for production that is not being used, such as plants and equipment idle for all or part of the day. Small increases in price make it profitable for firms to begin using this idle capacity. As the quantity supplied rises, firms begin to reach capacity. Once capacity is fully used, increasing production further requires the construction of new plants. To induce firms to incur this extra expense, the price must rise substantially, so supply becomes less elastic.

Figure 6 presents a numerical example of this phenomenon. When the price rises from $3 to $4 (a 29 percent increase, according to the midpoint method), the quantity supplied rises from 100 to 200 (a 67 percent increase). Because quantity supplied changes proportionately more than the price, the supply curve has elasticity greater than 1. By contrast, when the price rises from $12 to $15 (a 22 percent increase), the quantity supplied rises from 500 to 525 (a 5 percent increase). In this case, quantity supplied moves proportionately less than the price, so the elasticity is less than 1.

**Quick Quiz** Define the price elasticity of supply. Explain why the price elasticity of supply might be different in the long run and in the short run.

**Three Applications of Supply, Demand, and Elasticity**

Can good news for farming be bad news for farmers? Why did OPEC fail to keep the price of oil high? Does drug interdiction increase or decrease drug-related crime? At first, these questions might seem to have little in common. Yet all three...
questions are about markets, and all markets are subject to the forces of supply and demand. Here we apply the versatile tools of supply, demand, and elasticity to answer these seemingly complex questions.

**Can Good News for Farming Be Bad News for Farmers?**

Imagine yourself as a Kansas wheat farmer. Because you earn all your income from selling wheat, you devote much effort to making your land as productive as possible. You monitor weather and soil conditions, check your fields for pests and disease, and study the latest advances in farm technology. You know that the more wheat you grow, the more you will have to sell after the harvest, and the higher will be your income and your standard of living.

One day, Kansas State University announces a major discovery. Researchers in its agronomy department have devised a new hybrid of wheat that raises the amount farmers can produce from each acre of land by 20 percent. How should you react to this news? Does this discovery make you better off or worse off than you were before?

Recall from Chapter 4 that we answer such questions in three steps. First, we examine whether the supply or demand curve shifts. Second, we consider in which direction the curve shifts. Third, we use the supply-and-demand diagram to see how the market equilibrium changes.

In this case, the discovery of the new hybrid affects the supply curve. Because the hybrid increases the amount of wheat that can be produced on each acre of land, farmers are now willing to supply more wheat at any given price. In other words, the supply curve shifts to the right. The demand curve remains the same because consumers’ desire to buy wheat products at any given price is not affected by the introduction of a new hybrid. Figure 7 shows an example of such a change. When the supply curve shifts from $S_1$ to $S_2$, the quantity of wheat sold increases from 100 to 110, and the price of wheat falls from $3 to $2.

![Figure 7](image)

**An Increase in Supply in the Market for Wheat**

When an advance in farm technology increases the supply of wheat from $S_1$ to $S_2$, the price of wheat falls. Because the demand for wheat is inelastic, the increase in the quantity sold from 100 to 110 is proportionately smaller than the decrease in the price from $3 to $2. As a result, farmers’ total revenue falls from $300 ($3 \times 100)$ to $220 ($2 \times 110)$.
Does this discovery make farmers better off? As a first cut to answering this question, consider what happens to the total revenue received by farmers. Farmers’ total revenue is \( P \times Q \), the price of the wheat times the quantity sold. The discovery affects farmers in two conflicting ways. The hybrid allows farmers to produce more wheat (\( Q \) rises), but now each bushel of wheat sells for less (\( P \) falls).

Whether total revenue rises or falls depends on the elasticity of demand. In practice, the demand for basic foodstuffs such as wheat is usually inelastic because these items are relatively inexpensive and have few good substitutes. When the demand curve is inelastic, as it is in Figure 7, a decrease in price causes total revenue to fall. You can see this in the figure: The price of wheat falls substantially, whereas the quantity of wheat sold rises only slightly. Total revenue falls from $300 to $220. Thus, the discovery of the new hybrid lowers the total revenue that farmers receive from the sale of their crops.

If farmers are made worse off by the discovery of this new hybrid, one might wonder why they adopt it. The answer goes to the heart of how competitive markets work. Because each farmer is only a small part of the market for wheat, he or she takes the price of wheat as given. For any given price of wheat, it is better to use the new hybrid to produce and sell more wheat. Yet when all farmers do this, the supply of wheat increases, the price falls, and farmers are worse off.

Although this example may at first seem hypothetical, it helps to explain a major change in the U.S. economy over the past century. Two hundred years ago, most Americans lived on farms. Knowledge about farm methods was sufficiently primitive that most Americans had to be farmers to produce enough food to feed the nation’s population. Yet over time, advances in farm technology increased the amount of food that each farmer could produce. This increase in food supply, together with inelastic food demand, caused farm revenues to fall, which in turn encouraged people to leave farming.

A few numbers show the magnitude of this historic change. As recently as 1950, there were 10 million people working on farms in the United States, representing 17 percent of the labor force. Today, fewer than 3 million people work on farms, or 2 percent of the labor force. This change coincided with tremendous advances in farm productivity: Despite the 70 percent drop in the number of farmers, U.S. farms now produce more than twice the output of crops and livestock that they did in 1950.
This analysis of the market for farm products also helps to explain a seeming paradox of public policy: Certain farm programs try to help farmers by inducing them not to plant crops on all of their land. The purpose of these programs is to reduce the supply of farm products and thereby raise prices. With inelastic demand for their products, farmers as a group receive greater total revenue if they supply a smaller crop to the market. No single farmer would choose to leave his land fallow on his own because each takes the market price as given. But if all farmers do so together, each of them can be better off.

When analyzing the effects of farm technology or farm policy, it is important to keep in mind that what is good for farmers is not necessarily good for society as a whole. Improvement in farm technology can be bad for farmers because it makes farmers increasingly unnecessary, but it is surely good for consumers who pay less for food. Similarly, a policy aimed at reducing the supply of farm products may raise the incomes of farmers, but it does so at the expense of consumers.

**Why Did OPEC Fail to Keep the Price of Oil High?**

Many of the most disruptive events for the world’s economies over the past several decades have originated in the world market for oil. In the 1970s, members of the Organization of Petroleum Exporting Countries (OPEC) decided to raise the world price of oil to increase their incomes. These countries accomplished this goal by jointly reducing the amount of oil they supplied. From 1973 to 1974, the price of oil (adjusted for overall inflation) rose more than 50 percent. Then, a few years later, OPEC did the same thing again. From 1979 to 1981, the price of oil approximately doubled.

Yet OPEC found it difficult to maintain a high price. From 1982 to 1985, the price of oil steadily declined about 10 percent per year. Dissatisfaction and disarray soon prevailed among the OPEC countries. In 1986, cooperation among OPEC members completely broke down, and the price of oil plunged 45 percent. In 1990, the price of oil (adjusted for overall inflation) was back to where it began in 1970, and it stayed at that low level throughout most of the 1990s. (In the first decade of the 21st century, the price of oil rose again, but the main driving force was not OPEC supply restrictions but, rather, increased world demand, in part from a large and rapidly growing Chinese economy.)

This OPEC episode of the 1970s and 1980s shows how supply and demand can behave differently in the short run and in the long run. In the short run, both the supply and demand for oil are relatively inelastic. Supply is inelastic because the quantity of known oil reserves and the capacity for oil extraction cannot be changed quickly. Demand is inelastic because buying habits do not respond immediately to changes in price. Thus, as panel (a) of Figure 8 shows, the short-run supply and demand curves are steep. When the supply of oil shifts from $S_1$ to $S_2$, the price increase from $P_1$ to $P_2$ is large.

The situation is very different in the long run. Over long periods of time, producers of oil outside OPEC respond to high prices by increasing oil exploration and by building new extraction capacity. Consumers respond with greater conservation, for instance by replacing old inefficient cars with newer efficient ones. Thus, as panel (b) of Figure 8 shows, the long-run supply and demand curves are more elastic. In the long run, the shift in the supply curve from $S_1$ to $S_2$ causes a much smaller increase in the price.
This analysis shows why OPEC succeeded in maintaining a high price of oil only in the short run. When OPEC countries agreed to reduce their production of oil, they shifted the supply curve to the left. Even though each OPEC member sold less oil, the price rose by so much in the short run that OPEC incomes rose. By contrast, in the long run, supply and demand are relatively elastic, as in panel (b). In this case, the same size shift in the supply curve ($S_1$ to $S_2$) causes a smaller increase in the price. Thus, OPEC’s coordinated reduction in supply proved less profitable in the long run. The cartel learned that raising prices is easier in the short run than in the long run.

**Does Drug Interdiction Increase or Decrease Drug-Related Crime?**

A persistent problem facing our society is the use of illegal drugs, such as heroin, cocaine, ecstasy, and crack. Drug use has several adverse effects. One is that drug dependence can ruin the lives of drug users and their families. Another is that drug addicts often turn to robbery and other violent crimes to obtain the money needed to support their habit. To discourage the use of illegal drugs, the U.S. government devotes billions of dollars each year to reduce the flow of drugs into the country. Let’s use the tools of supply and demand to examine this policy of drug interdiction.
Suppose the government increases the number of federal agents devoted to the war on drugs. What happens in the market for illegal drugs? As is usual, we answer this question in three steps. First, we consider whether the supply or demand curve shifts. Second, we consider the direction of the shift. Third, we see how the shift affects the equilibrium price and quantity.

Although the purpose of drug interdiction is to reduce drug use, its direct impact is on the sellers of drugs rather than the buyers. When the government stops some drugs from entering the country and arrests more smugglers, it raises the cost of selling drugs and, therefore, reduces the quantity of drugs supplied at any given price. The demand for drugs—the amount buyers want at any given price—is not changed. As panel (a) of Figure 9 shows, interdiction shifts the supply curve to the left from $S_1$ to $S_2$ and leaves the demand curve the same. The equilibrium price of drugs rises from $P_1$ to $P_2$, and the equilibrium quantity falls from $Q_1$ to $Q_2$. The fall in the equilibrium quantity shows that drug interdiction does reduce drug use.

But what about the amount of drug-related crime? To answer this question, consider the total amount that drug users pay for the drugs they buy. Because few drug addicts are likely to break their destructive habits in response to a higher price, it is likely that the demand for drugs is inelastic, as it is drawn in the figure.

Drug interdiction reduces the supply of drugs from $S_1$ to $S_2$, as in panel (a). If the demand for drugs is inelastic, then the total amount paid by drug users rises, even as the amount of drug use falls. By contrast, drug education reduces the demand for drugs from $D_1$ to $D_2$, as in panel (b). Because both price and quantity fall, the amount paid by drug users falls.
If demand is inelastic, then an increase in price raises total revenue in the drug market. That is, because drug interdiction raises the price of drugs proportionately more than it reduces drug use, it raises the total amount of money that drug users pay for drugs. Addicts who already had to steal to support their habits would have an even greater need for quick cash. Thus, drug interdiction could increase drug-related crime.

Because of this adverse effect of drug interdiction, some analysts argue for alternative approaches to the drug problem. Rather than trying to reduce the supply of drugs, policymakers might try to reduce the demand by pursuing a policy of drug education. Successful drug education has the effects shown in panel (b) of Figure 9. The demand curve shifts to the left from $D_1$ to $D_2$. As a result, the equilibrium quantity falls from $Q_1$ to $Q_2$, and the equilibrium price falls from $P_1$ to $P_2$. Total revenue, which is price times quantity, also falls. Thus, in contrast to drug interdiction, drug education can reduce both drug use and drug-related crime.

Advocates of drug interdiction might argue that the long-run effects of this policy are different from the short-run effects because the elasticity of demand depends on the time horizon. The demand for drugs is probably inelastic over short periods because higher prices do not substantially affect drug use by established addicts. But demand may be more elastic over longer periods because higher prices would discourage experimentation with drugs among the young and, over time, lead to fewer drug addicts. In this case, drug interdiction would increase drug-related crime in the short run while decreasing it in the long run.

**Quick Quiz** How might a drought that destroys half of all farm crops be good for farmers? If such a drought is good for farmers, why don’t farmers destroy their own crops in the absence of a drought?

**Conclusion**

According to an old quip, even a parrot can become an economist simply by learning to say “supply and demand.” These last two chapters should have convinced you that there is much truth in this statement. The tools of supply and demand allow you to analyze many of the most important events and policies that shape the economy. You are now well on your way to becoming an economist (or at least a well-educated parrot).

**Summary**

- The price elasticity of demand measures how much the quantity demanded responds to changes in the price. Demand tends to be more elastic if close substitutes are available, if the good is a luxury rather than a necessity, if the market is narrowly defined, or if buyers have substantial time to react to a price change.

- The price elasticity of demand is calculated as the percentage change in quantity demanded divided by the percentage change in price. If quantity demanded moves proportionately less than the price, then the elasticity is less than 1, and demand is said to be inelastic. If quantity demanded moves proportionately more than the
price, then the elasticity is greater than 1, and demand is said to be elastic.

- Total revenue, the total amount paid for a good, equals the price of the good times the quantity sold. For inelastic demand curves, total revenue rises as price rises. For elastic demand curves, total revenue falls as price rises.

- The income elasticity of demand measures how much the quantity demanded responds to changes in consumers’ income. The cross-price elasticity of demand measures how much the quantity demanded of one good responds to changes in the price of another good.

- The price elasticity of supply measures how much the quantity supplied responds to changes in the price. This elasticity often depends on the time horizon under consideration. In most markets, supply is more elastic in the long run than in the short run.

- The price elasticity of supply is calculated as the percentage change in quantity supplied divided by the percentage change in price. If quantity supplied moves proportionately less than the price, then the elasticity is less than 1, and supply is said to be inelastic. If quantity supplied moves proportionately more than the price, then the elasticity is greater than 1, and supply is said to be elastic.

- The tools of supply and demand can be applied in many different kinds of markets. This chapter uses them to analyze the market for wheat, the market for oil, and the market for illegal drugs.

**KEY CONCEPTS**

- elasticity, p. 90
- price elasticity of demand, p. 90
- total revenue, p. 93
- income elasticity of demand, p. 97
- cross-price elasticity of demand, p. 99
- price elasticity of supply, p. 99

**QUESTIONS FOR REVIEW**

1. Define the price elasticity of demand and the income elasticity of demand.
2. List and explain the four determinants of the price elasticity of demand discussed in the chapter.
3. What is the main advantage of using the midpoint method for calculating elasticity?
4. If the elasticity is greater than 1, is demand elastic or inelastic? If the elasticity equals 0, is demand perfectly elastic or perfectly inelastic?
5. On a supply-and-demand diagram, show equilibrium price, equilibrium quantity, and the total revenue received by producers.
6. If demand is elastic, how will an increase in price change total revenue? Explain.
7. What do we call a good whose income elasticity is less than 0?
8. How is the price elasticity of supply calculated? Explain what it measures.
9. What is the price elasticity of supply of Picasso paintings?
10. Is the price elasticity of supply usually larger in the short run or in the long run? Why?
11. How did elasticity help explain why drug interdiction could reduce the supply of drugs, yet possibly increase drug-related crime?
1. For each of the following pairs of goods, which good would you expect to have more elastic demand and why?
   a. required textbooks or mystery novels
   b. Beethoven recordings or classical music recordings in general
   c. subway rides during the next 6 months or subway rides during the next 5 years
   d. root beer or water

2. Suppose that business travelers and vacationers have the following demand for airline tickets from New York to Boston:

<table>
<thead>
<tr>
<th>Price</th>
<th>Quantity Demanded (business travelers)</th>
<th>Quantity Demanded (vacationers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$150</td>
<td>2,100 tickets</td>
<td>1,000 tickets</td>
</tr>
<tr>
<td>200</td>
<td>2,000</td>
<td>800</td>
</tr>
<tr>
<td>250</td>
<td>1,900</td>
<td>600</td>
</tr>
<tr>
<td>300</td>
<td>1,800</td>
<td>400</td>
</tr>
</tbody>
</table>

   a. As the price of tickets rises from $200 to $250, what is the price elasticity of demand for (i) business travelers and (ii) vacationers? (Use the midpoint method in your calculations.)
   b. Why might vacationers have a different elasticity from business travelers?

3. Suppose the price elasticity of demand for heating oil is 0.2 in the short run and 0.7 in the long run.
   a. If the price of heating oil rises from $1.80 to $2.20 per gallon, what happens to the quantity of heating oil demanded in the short run? In the long run? (Use the midpoint method in your calculations.)
   b. Why might this elasticity depend on the time horizon?

4. A price change causes the quantity demanded of a good to decrease by 30 percent, while the total revenue of that good increases by 15 percent. Is the demand curve elastic or inelastic? Explain.

5. The equilibrium price of coffee mugs rose sharply last month, but the equilibrium quantity was the same as ever. Three people tried to explain the situation. Which explanations could be right? Explain your logic.

6. Suppose that your demand schedule for compact discs is as follows:

<table>
<thead>
<tr>
<th>Price</th>
<th>Quantity Demanded (income = $10,000)</th>
<th>Quantity Demanded (income = $12,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ 8</td>
<td>40 CDs</td>
<td>50 CDs</td>
</tr>
<tr>
<td>10</td>
<td>32</td>
<td>45</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>30</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>12</td>
</tr>
</tbody>
</table>

   a. Use the midpoint method to calculate your price elasticity of demand as the price of compact discs increases from $8 to $10 if (i) your income is $10,000 and (ii) your income is $12,000.
   b. Calculate your income elasticity of demand as your income increases from $10,000 to $12,000 if (i) the price is $12 and (ii) the price is $16.

7. You have the following information about good X and good Y:
   • Income elasticity of demand for good X: −3
   • Cross-price elasticity of demand for good X with respect to the price of good Y: 2
   Would an increase in income and a decrease in the price of good Y unambiguously decrease the demand for good X? Why or why not?

8. Maria has decided always to spend one-third of her income on clothing.
   a. What is her income elasticity of clothing demand?
   b. What is her price elasticity of clothing demand?
   c. If Maria’s tastes change and she decides to spend only one-fourth of her income on clothing, how does her demand curve change? What is her income elasticity and price elasticity now?
9. The *New York Times* reported (Feb. 17, 1996, p. 25) that subway ridership declined after a fare increase: “There were nearly four million fewer riders in December 1995, the first full month after the price of a token increased 25 cents to $1.50, than in the previous December, a 4.3 percent decline.”
   a. Use these data to estimate the price elasticity of demand for subway rides.
   b. According to your estimate, what happens to the Transit Authority’s revenue when the fare rises?
   c. Why might your estimate of the elasticity be unreliable?
10. Two drivers—Tom and Jerry—each drive up to a gas station. Before looking at the price, each places an order. Tom says, “I’d like 10 gallons of gas.” Jerry says, “I’d like $10 worth of gas.”
    What is each driver’s price elasticity of demand?
11. Consider public policy aimed at smoking.
    a. Studies indicate that the price elasticity of demand for cigarettes is about 0.4. If a pack of cigarettes currently costs $2 and the government wants to reduce smoking by 20 percent, by how much should it increase the price?
    b. If the government permanently increases the price of cigarettes, will the policy have a larger effect on smoking 1 year from now or 5 years from now?
    c. Studies also find that teenagers have a higher price elasticity than do adults. Why might this be true?
12. You are the curator of a museum. The museum is running short of funds, so you decide to increase revenue. Should you increase or decrease the price of admission? Explain.
13. Pharmaceutical drugs have an inelastic demand, and computers have an elastic demand. Suppose that technological advance doubles the supply of both products (that is, the quantity supplied at each price is twice what it was).
   a. What happens to the equilibrium price and quantity in each market?
   b. Which product experiences a larger change in price?
   c. Which product experiences a larger change in quantity?
   d. What happens to total consumer spending on each product?
14. Beachfront resorts have an inelastic supply, and automobiles have an elastic supply. Suppose that a rise in population doubles the demand for both products (that is, the quantity demanded at each price is twice what it was).
   a. What happens to the equilibrium price and quantity in each market?
   b. Which product experiences a larger change in price?
   c. Which product experiences a larger change in quantity?
   d. What happens to total consumer spending on each product?
15. Several years ago, flooding along the Missouri and the Mississippi rivers destroyed thousands of acres of wheat.
   a. Farmers whose crops were destroyed by the floods were much worse off, but farmers whose crops were not destroyed benefited from the floods. Why?
   b. What information would you need about the market for wheat to assess whether farmers as a group were hurt or helped by the floods?
16. Explain why the following might be true: A drought around the world raises the total revenue that farmers receive from the sale of grain, but a drought only in Kansas reduces the total revenue that Kansas farmers receive.
17. Suppose the demand curve for a product is \( Q = \frac{60}{P} \). Compute the quantity demanded at prices of $1, $2, $3, $4, $5, and $6. Graph the demand curve. Use the midpoint method to calculate the price elasticity of demand between $1 and $2 and between $5 and $6. How does this demand curve compare to the linear demand curve?
This page intentionally left blank
Supply, Demand, and Government Policies

Economists have two roles. As scientists, they develop and test theories to explain the world around them. As policy advisers, they use their theories to help change the world for the better. The focus of the preceding two chapters has been scientific. We have seen how supply and demand determine the price of a good and the quantity of the good sold. We have also seen how various events shift supply and demand and thereby change the equilibrium price and quantity.

This chapter offers our first look at policy. Here we analyze various types of government policy using only the tools of supply and demand. As you will see, the analysis yields some surprising insights. Policies often have effects that their architects did not intend or anticipate.

We begin by considering policies that directly control prices. For example, rent-control laws dictate a maximum rent that landlords may charge tenants. Minimum-wage laws dictate the lowest wage that firms may pay workers. Price controls are usually enacted when policymakers believe that the market price of a good or service is unfair to buyers or sellers. Yet, as we will see, these policies can generate inequities of their own.
CONTROLS ON PRICES

To see how price controls affect market outcomes, let’s look once again at the market for ice cream. As we saw in Chapter 4, if ice cream is sold in a competitive market free of government regulation, the price of ice cream adjusts to balance supply and demand: At the equilibrium price, the quantity of ice cream that buyers want to buy exactly equals the quantity that sellers want to sell. To be concrete, suppose the equilibrium price is $3 per cone.

Not everyone may be happy with the outcome of this free-market process. Let’s say the American Association of Ice-Cream Eaters complains that the $3 price is too high for everyone to enjoy a cone a day (their recommended diet). Meanwhile, the National Organization of Ice-Cream Makers complains that the $3 price—the result of “cutthroat competition”—is too low and is depressing the incomes of its members. Each of these groups lobbies the government to pass laws that alter the market outcome by directly controlling the price of an ice-cream cone.

Because buyers of any good always want a lower price while sellers want a higher price, the interests of the two groups conflict. If the Ice-Cream Eaters are successful in their lobbying, the government imposes a price ceiling on the market for ice cream. Because the price that balances supply and demand ($3) is below the ceiling, the legislated maximum is called a price ceiling. By contrast, if the Ice-Cream Makers are successful, the government imposes a legal minimum on the price. Because the price cannot fall below this level, the legislated minimum is called a price floor. Let us consider the effects of these policies in turn.

HOW PRICE CEILINGS AFFECT MARKET OUTCOMES

When the government, moved by the complaints and campaign contributions of the Ice-Cream Eaters, imposes a price ceiling on the market for ice cream, two outcomes are possible. In panel (a) of Figure 1, the government imposes a price ceiling of $4 per cone. In this case, because the price that balances supply and demand ($3) is below the ceiling, the price ceiling is not binding. Market forces naturally move the economy to the equilibrium, and the price ceiling has no effect on the price or the quantity sold.

Panel (b) of Figure 1 shows the other, more interesting, possibility. In this case, the government imposes a price ceiling of $2 per cone. Because the equilibrium price of $3 is above the price ceiling, the ceiling is a binding constraint on the market. The forces of supply and demand tend to move the price toward the equilibrium price, but when the market price hits the ceiling, it can, by law, rise no further. Thus, the market price equals the price ceiling. At this price, the quantity of ice cream demanded (125 cones in the figure) exceeds the quantity supplied (75 cones). There is a shortage of ice cream: 50 people who want to buy ice cream at the going price are unable to do so.
When a shortage of ice cream develops because of this price ceiling, some mechanism for rationing ice cream will naturally develop. The mechanism could be long lines: Buyers who are willing to arrive early and wait in line get a cone, but those unwilling to wait do not. Alternatively, sellers could ration ice cream according to their own personal biases, selling it only to friends, relatives, or members of their own racial or ethnic group. Notice that even though the price ceiling was motivated by a desire to help buyers of ice cream, not all buyers benefit from the policy. Some buyers do get to pay a lower price, although they may have to wait in line to do so, but other buyers cannot get any ice cream at all.

This example in the market for ice cream shows a general result: When the government imposes a binding price ceiling on a competitive market, a shortage of the good arises, and sellers must ration the scarce goods among the large number of potential buyers. The rationing mechanisms that develop under price ceilings are rarely desirable. Long lines are inefficient because they waste buyers’ time. Discrimination according to seller bias is both inefficient (because the good does not necessarily go to the buyer who values it most highly) and potentially unfair. By contrast, the rationing mechanism in a free, competitive market is both efficient and impersonal. When the market for ice cream reaches its equilibrium, anyone who wants to pay the market price can get a cone. Free markets ration goods with prices.
LINES AT THE GAS PUMP

As we discussed in the preceding chapter, in 1973 the Organization of Petroleum Exporting Countries (OPEC) raised the price of crude oil in world oil markets. Because crude oil is the major input used to make gasoline, the higher oil prices reduced the supply of gasoline. Long lines at gas stations became commonplace, and motorists often had to wait for hours to buy only a few gallons of gas.

What was responsible for the long gas lines? Most people blame OPEC. Surely, if OPEC had not raised the price of crude oil, the shortage of gasoline would not have occurred. Yet economists blame U.S. government regulations that limited the price oil companies could charge for gasoline.

Figure 2 shows what happened. As shown in panel (a), before OPEC raised the price of crude oil, the equilibrium price of gasoline, $P_1$, was below the price ceiling. The price regulation, therefore, had no effect. When the price of crude oil rose, however, the situation changed. The increase in the price of crude oil raised the cost of producing gasoline, and this reduced the supply of gasoline. As panel (b) shows, the supply curve shifted to the left from $S_1$ to $S_2$. In an unregulated market, this shift in supply would have raised the equilibrium price of gasoline from $P_1$ to $P_2$, and no shortage would have resulted. Instead, the price ceiling prevented the price from rising to the equilibrium level. At the price ceiling, producers were willing to sell $Q_S$, and consumers were willing to buy $Q_D$. Thus, the shift in supply caused a severe shortage at the regulated price.

Panel (a) shows the gasoline market when the price ceiling is not binding because the equilibrium price, $P_1$, is below the ceiling. Panel (b) shows the gasoline market after an increase in the price of crude oil (an input into making gasoline) shifts the supply curve to the left from $S_1$ to $S_2$. In an unregulated market, the price would have risen from $P_1$ to $P_2$. The price ceiling, however, prevents this from happening. At the binding price ceiling, consumers are willing to buy $Q_D$, but producers of gasoline are willing to sell only $Q_S$. The difference between quantity demanded and quantity supplied, $Q_D - Q_S$, measures the gasoline shortage.
Eventually, the laws regulating the price of gasoline were repealed. Lawmakers came to understand that they were partly responsible for the many hours Americans lost waiting in line to buy gasoline. Today, when the price of crude oil changes, the price of gasoline can adjust to bring supply and demand into equilibrium.

**RENT CONTROL IN THE SHORT RUN AND THE LONG RUN**

One common example of a price ceiling is rent control. In many cities, the local government places a ceiling on rents that landlords may charge their tenants. The goal of this policy is to help the poor by making housing more affordable. Economists often criticize rent control, arguing that it is a highly inefficient way to help the poor raise their standard of living. One economist called rent control “the best way to destroy a city, other than bombing.”

The adverse effects of rent control are less apparent to the general population because these effects occur over many years. In the short run, landlords have a fixed number of apartments to rent, and they cannot adjust this number quickly as market conditions change. Moreover, the number of people searching for housing in a city may not be highly responsive to rents in the short run because people take time to adjust their housing arrangements. Therefore, the short-run supply and demand for housing are relatively inelastic.

Panel (a) of Figure 3 shows the short-run effects of rent control on the housing market. As with any binding price ceiling, rent control causes a shortage. Yet
because supply and demand are inelastic in the short run, the initial shortage caused by rent control is small. The primary effect in the short run is to reduce rents.

The long-run story is very different because the buyers and sellers of rental housing respond more to market conditions as time passes. On the supply side, landlords respond to low rents by not building new apartments and by failing to maintain existing ones. On the demand side, low rents encourage people to find their own apartments (rather than living with their parents or sharing apartments with roommates) and induce more people to move into a city. Therefore, both supply and demand are more elastic in the long run.

Panel (b) of Figure 3 illustrates the housing market in the long run. When rent control depresses rents below the equilibrium level, the quantity of apartments supplied falls substantially, and the quantity of apartments demanded rises substantially. The result is a large shortage of housing.

In cities with rent control, landlords use various mechanisms to ration housing. Some landlords keep long waiting lists. Others give a preference to tenants without children. Still others discriminate on the basis of race. Sometimes apartments are allocated to those willing to offer under-the-table payments to building superintendents. In essence, these bribes bring the total price of an apartment (including the bribe) closer to the equilibrium price.

To understand fully the effects of rent control, we have to remember one of the Ten Principles of Economics from Chapter 1: People respond to incentives. In free markets, landlords try to keep their buildings clean and safe because desirable apartments command higher prices. By contrast, when rent control creates shortages and waiting lists, landlords lose their incentive to respond to tenants’ concerns. Why should a landlord spend money to maintain and improve the property when people are waiting to get in as it is? In the end, tenants get lower rents, but they also get lower-quality housing.

Policymakers often react to the effects of rent control by imposing additional regulations. For example, there are laws that make racial discrimination in housing illegal and require landlords to provide minimally adequate living conditions. These laws, however, are difficult and costly to enforce. By contrast, when rent control is eliminated and a market for housing is regulated by the forces of competition, such laws are less necessary. In a free market, the price of housing adjusts to eliminate the shortages that give rise to undesirable landlord behavior.

**How Price Floors Affect Market Outcomes**

To examine the effects of another kind of government price control, let’s return to the market for ice cream. Imagine now that the government is persuaded by the pleas of the National Organization of Ice-Cream Makers. In this case, the government might institute a price floor. Price floors, like price ceilings, are an attempt by the government to maintain prices at other than equilibrium levels. Whereas a price ceiling places a legal maximum on prices, a price floor places a legal minimum.

When the government imposes a price floor on the ice-cream market, two outcomes are possible. If the government imposes a price floor of $2 per cone when the equilibrium price is $3, we obtain the outcome in panel (a) of Figure 4. In this case, because the equilibrium price is above the floor, the price floor is not binding. Market forces naturally move the economy to the equilibrium, and the price floor has no effect.
In panel (a), the government imposes a price floor of $2. Because this is below the equilibrium price of $3, the price floor has no effect. The market price adjusts to balance supply and demand. At the equilibrium, quantity supplied and quantity demanded both equal 100 cones. In panel (b), the government imposes a price floor of $4, which is above the equilibrium price of $3. Therefore, the market price equals $4. Because 120 cones are supplied at this price and only 80 are demanded, there is a surplus of 40 cones.

Panel (b) of Figure 4 shows what happens when the government imposes a price floor of $4 per cone. In this case, because the equilibrium price of $3 is below the floor, the price floor is a binding constraint on the market. The forces of supply and demand tend to move the price toward the equilibrium price, but when the market price hits the floor, it can fall no further. The market price equals the price floor. At this floor, the quantity of ice cream supplied (120 cones) exceeds the quantity demanded (80 cones). Some people who want to sell ice cream at the going price are unable to. Thus, a binding price floor causes a surplus.

Just as the shortages resulting from price ceilings can lead to undesirable rationing mechanisms, so can the surpluses resulting from price floors. In the case of a price floor, some sellers are unable to sell all they want at the market price. The sellers who appeal to the personal biases of the buyers, perhaps due to racial or familial ties, are better able to sell their goods than those who do not. By contrast, in a free market, the price serves as the rationing mechanism, and sellers can sell all they want at the equilibrium price.

**THE MINIMUM WAGE**

An important example of a price floor is the minimum wage. Minimum-wage laws dictate the lowest price for labor that any employer may pay. The U.S. Congress first instituted a minimum wage with the Fair Labor Standards Act of 1938 to ensure workers a minimally adequate standard of living. In 2007, the minimum
wage according to federal law was $5.15 per hour, and it was scheduled to increase to $7.25 by 2010. (Some states mandate minimum wages above the federal level.) Most European nations have minimum-wage laws as well; some, such as France and the United Kingdom, have significantly higher minimums than the United States.

To examine the effects of a minimum wage, we must consider the market for labor. Panel (a) of Figure 5 shows the labor market, which, like all markets, is subject to the forces of supply and demand. Workers determine the supply of labor, and firms determine the demand. If the government doesn’t intervene, the wage normally adjusts to balance labor supply and labor demand.

Panel (b) of Figure 5 shows the labor market with a minimum wage. If the minimum wage is above the equilibrium level, as it is here, the quantity of labor supplied exceeds the quantity demanded. The result is unemployment. Thus, the minimum wage raises the incomes of those workers who have jobs, but it lowers the incomes of workers who cannot find jobs.

To fully understand the minimum wage, keep in mind that the economy contains not a single labor market but many labor markets for different types of workers. The impact of the minimum wage depends on the skill and experience of the worker. Workers with high skills and much experience are not affected because their equilibrium wages are well above the minimum. For these workers, the minimum wage is not binding.

The minimum wage has its greatest impact on the market for teenage labor. The equilibrium wages of teenagers are low because teenagers are among the least skilled and least experienced members of the labor force. In addition, teenagers are often willing to accept a lower wage in exchange for on-the-job training. (Some

---

**FIGURE**

How the Minimum Wage Affects the Labor Market

Panel (a) shows a labor market in which the wage adjusts to balance labor supply and labor demand. Panel (b) shows the impact of a binding minimum wage. Because the minimum wage is a price floor, it causes a surplus: The quantity of labor supplied exceeds the quantity demanded. The result is unemployment.
teenagers are willing to work as “interns” for no pay at all. Because internships pay nothing, however, the minimum wage does not apply to them. If it did, these jobs might not exist.) As a result, the minimum wage is more often binding for teenagers than for other members of the labor force.

Many economists have studied how minimum-wage laws affect the teenage labor market. These researchers compare the changes in the minimum wage over time with the changes in teenage employment. Although there is some debate about how much the minimum wage affects employment, the typical study finds that a 10 percent increase in the minimum wage depresses teenage employment between 1 and 3 percent. In interpreting this estimate, note that a 10 percent increase in the minimum wage does not raise the average wage of teenagers by 10 percent. A change in the law does not directly affect those teenagers who are already paid well above the minimum, and enforcement of minimum-wage laws is not perfect. Thus, the estimated drop in employment of 1 to 3 percent is significant.

In addition to altering the quantity of labor demanded, the minimum wage alters the quantity supplied. Because the minimum wage raises the wage that teenagers can earn, it increases the number of teenagers who choose to look for jobs. Studies have found that a higher minimum wage influences which teenagers are employed. When the minimum wage rises, some teenagers who are still attending school choose to drop out and take jobs. These new dropouts displace other teenagers who had already dropped out of school and who now become unemployed.

The minimum wage is a frequent topic of debate. Economists are about evenly divided on the issue. In a 2006 survey of PhD economists, 47 percent favored eliminating the minimum wage, while 14 percent would maintain it at its current level and 38 percent would increase it.

Advocates of the minimum wage view the policy as one way to raise the income of the working poor. They correctly point out that workers who earn the minimum wage can afford only a meager standard of living. In 2007, for instance, when the minimum wage was $5.15 per hour, two adults working 40 hours a week for every week of the year at minimum-wage jobs had a total annual income of only $21,424, which was less than half of the median family income. Many advocates of the minimum wage admit that it has some adverse effects, including unemployment, but they believe that these effects are small and that, all things considered, a higher minimum wage makes the poor better off.

Opponents of the minimum wage contend that it is not the best way to combat poverty. They note that a high minimum wage causes unemployment, encourages teenagers to drop out of school, and prevents some unskilled workers from getting the on-the-job training they need. Moreover, opponents of the minimum wage point out that it is a poorly targeted policy. Not all minimum-wage workers are heads of households trying to help their families escape poverty. In fact, fewer than a third of minimum-wage earners are in families with incomes below the poverty line. Many are teenagers from middle-class homes working at part-time jobs for extra spending money.

**Evaluating Price Controls**

One of the Ten Principles of Economics discussed in Chapter 1 is that markets are usually a good way to organize economic activity. This principle explains why economists usually oppose price ceilings and price floors. To economists, prices
President Chavez versus the Market

Venezuela’s president has tried to replace market prices with his own.

Price Caps Ail Venezuelan Economy

By Peter Millard and Raul Gallegos

CARACAS, VENEZUELA—After 21 years in the milk business, Ismael Cárdenas Gil is throwing in the towel.

Mr. Cárdenas, who heads Alimentaria Internacional, can no longer make a profit selling imported powdered milk under government-imposed price controls. As a result, he has cut back his imports to “practically zero.”

“The controls have been very harsh. The numbers don’t work out to import milk and sell it here,” Mr. Cárdenas says.

His plight is becoming more common in Venezuela, with President Hugo Chávez meddling in the economy to advance his populist-leftist agenda as companies selling price-regulated products watch their profits disappear. While government controls have slowed the growth of inflation, Venezuela’s rate is still the highest in Latin America. The controls also have led to frequent product shortages and the emergence of a thriving black market. Some farmers and retailers are skirting the rules or have stopped selling certain goods altogether rather than sell them at a loss.

The problems facing Venezuelan businesses and consumers serve as a cautionary tale for the growing ranks of Latin American populists pushing for a heavy government hand in the economy. Mr. Chávez is taking advantage of the country’s massive oil-revenue windfall to fund a governing philosophy he has dubbed “socialism for the 21st century.” His goal is to increase social spending and curb inflation through a mix of price caps, a fixed exchange rate and fixed interest rates.

But some Venezuelan businesses hurt by the price controls are beginning to balk. Last week, corn growers marched outside the presidential palace, protesting government controls they say have dried up demand for their corn. While the farmers are getting a
shortages. Similarly, wage subsidies raise the living standards of the working poor without discouraging firms from hiring them. An example of a wage subsidy is the earned income tax credit, a government program that supplements the incomes of low-wage workers.

Although these alternative policies are often better than price controls, they are not perfect. Rent and wage subsidies cost the government money and, therefore, require higher taxes. As we see in the next section, taxation has costs of its own.

**Quick Quiz** Define *price ceiling* and *price floor* and give an example of each. Which leads to a shortage? Which leads to a surplus? Why?

**Taxes**

All governments—from the federal government in Washington, D.C., to the local governments in small towns—use taxes to raise revenue for public projects, such as roads, schools, and national defense. Because taxes are such an important policy instrument, and because they affect our lives in many ways, we return to the
study of taxes several times throughout this book. In this section, we begin our study of how taxes affect the economy.

To set the stage for our analysis, imagine that a local government decides to hold an annual ice-cream celebration—with a parade, fireworks, and speeches by town officials. To raise revenue to pay for the event, the town decides to place a $0.50 tax on the sale of ice-cream cones. When the plan is announced, our two lobbying groups swing into action. The American Association of Ice-Cream Eaters claims that consumers of ice cream are having trouble making ends meet, and it argues that sellers of ice cream should pay the tax. The National Organization of Ice-Cream Makers claims that its members are struggling to survive in a competitive market, and it argues that buyers of ice cream should pay the tax. The town mayor, hoping to reach a compromise, suggests that half the tax be paid by the buyers and half be paid by the sellers.

To analyze these proposals, we need to address a simple but subtle question: When the government levies a tax on a good, who actually bears the burden of the tax? The people buying the good? The people selling the good? Or if buyers and sellers share the tax burden, what determines how the burden is divided? Can the government simply legislate the division of the burden, as the mayor is suggesting, or is the division determined by more fundamental market forces? The term tax incidence refers to how the burden of a tax is distributed among the various people who make up the economy. As we will see, some surprising lessons about tax incidence can be learned by applying the tools of supply and demand.

**HOW TAXES ON SELLERS AFFECT MARKET OUTCOMES**

We begin by considering a tax levied on sellers of a good. Suppose the local government passes a law requiring sellers of ice-cream cones to send $0.50 to the government for each cone they sell. How does this law affect the buyers and sellers of ice cream? To answer this question, we can follow the three steps in Chapter 4 for analyzing supply and demand: (1) We decide whether the law affects the supply curve or demand curve. (2) We decide which way the curve shifts. (3) We examine how the shift affects the equilibrium price and quantity.

**Step One** The immediate impact of the tax is on the sellers of ice cream. Because the tax is not levied on buyers, the quantity of ice cream demanded at any given price is the same; thus, the demand curve does not change. By contrast, the tax on sellers makes the ice-cream business less profitable at any given price, so it shifts the supply curve.

**Step Two** Because the tax on sellers raises the cost of producing and selling ice cream, it reduces the quantity supplied at every price. The supply curve shifts to the left (or, equivalently, upward).

We can, in this case, be precise about how much the curve shifts. For any market price of ice cream, the effective price to sellers—the amount they get to keep after paying the tax—is $0.50 lower. For example, if the market price of a cone happened to be $2.00, the effective price received by sellers would be $1.50. Whatever the market price, sellers will supply a quantity of ice cream as if the price were $0.50 lower than it is. Put differently, to induce sellers to supply any given quantity, the market price must now be $0.50 higher to compensate for the effect of the tax. Thus, as shown in Figure 6, the supply curve shifts upward from $S_1$ to $S_2$ by the exact size of the tax ($0.50).
Step Three  Having determined how the supply curve shifts, we can now compare the initial and the new equilibriums. The figure shows that the equilibrium price of ice cream rises from $3.00 to $3.30, and the equilibrium quantity falls from 100 to 90 cones. Because sellers sell less and buyers buy less in the new equilibrium, the tax reduces the size of the ice-cream market.

Implications  We can now return to the question of tax incidence: Who pays the tax? Although sellers send the entire tax to the government, buyers and sellers share the burden. Because the market price rises from $3.00 to $3.30 when the tax is introduced, buyers pay $0.30 more for each ice-cream cone than they did without the tax. Thus, the tax makes buyers worse off. Sellers get a higher price ($3.30) from buyers than they did previously, but the effective price after paying the tax falls from $3.00 before the tax to $2.80 with the tax ($3.30 − $0.50 = $2.80). Thus, the tax also makes sellers worse off.

To sum up, the analysis yields two lessons:

- Taxes discourage market activity. When a good is taxed, the quantity of the good sold is smaller in the new equilibrium.
- Buyers and sellers share the burden of taxes. In the new equilibrium, buyers pay more for the good, and sellers receive less.

HOW TAXES ON BUYERS AFFECT MARKET OUTCOMES

Now consider a tax levied on buyers of a good. Suppose that our local government passes a law requiring buyers of ice-cream cones to send $0.50 to the government for each ice-cream cone they buy. What are the effects of this law? Again, we apply our three steps.
Step One The initial impact of the tax is on the demand for ice cream. The supply curve is not affected because, for any given price of ice cream, sellers have the same incentive to provide ice cream to the market. By contrast, buyers now have to pay a tax to the government (as well as the price to the sellers) whenever they buy ice cream. Thus, the tax shifts the demand curve for ice cream.

Step Two We next determine the direction of the shift. Because the tax on buyers makes buying ice cream less attractive, buyers demand a smaller quantity of ice cream at every price. As a result, the demand curve shifts to the left (or, equivalently, downward), as shown in Figure 7.

Once again, we can be precise about the magnitude of the shift. Because of the $0.50 tax levied on buyers, the effective price to buyers is now $0.50 higher than the market price (whatever the market price happens to be). For example, if the market price of a cone happened to be $2.00, the effective price to buyers would be $2.50. Because buyers look at their total cost including the tax, they demand a quantity of ice cream as if the market price were $0.50 higher than it actually is. In other words, to induce buyers to demand any given quantity, the market price must now be $0.50 lower to make up for the effect of the tax. Thus, the tax shifts the demand curve downward from $D_1$ to $D_2$ by the exact size of the tax ($0.50$).

Step Three Having determined how the demand curve shifts, we can now see the effect of the tax by comparing the initial equilibrium and the new equilibrium. You can see in the figure that the equilibrium price of ice cream falls from $3.00 to $2.80 and the equilibrium quantity falls from 100 to 90 cones. Once again, the tax on ice cream reduces the size of the ice-cream market. And once again, buyers and sellers share the burden of the tax. Sellers get a lower price for their product; buyers pay a lower market price to sellers than they did previously, but the effective price (including the tax buyers have to pay) rises from $3.00 to $3.30.

A Tax on Buyers
When a tax of $0.50 is levied on buyers, the demand curve shifts down by $0.50 from $D_1$ to $D_2$. The equilibrium quantity falls from 100 to 90 cones. The price that sellers receive falls from $3.00 to $2.80. The price that buyers pay (including the tax) rises from $3.00 to $3.30. Even though the tax is levied on buyers, buyers and sellers share the burden of the tax.

A tax on buyers shifts the demand curve downward by the size of the tax ($0.50$).
Implications If you compare Figures 6 and 7, you will notice a surprising conclusion: *Taxes levied on sellers and taxes levied on buyers are equivalent.* In both cases, the tax places a wedge between the price that buyers pay and the price that sellers receive. The wedge between the buyers’ price and the sellers’ price is the same, regardless of whether the tax is levied on buyers or sellers. In either case, the wedge shifts the relative position of the supply and demand curves. In the new equilibrium, buyers and sellers share the burden of the tax. The only difference between taxes on sellers and taxes on buyers is who sends the money to the government.

The equivalence of these two taxes is easy to understand if we imagine that the government collects the $0.50 ice-cream tax in a bowl on the counter of each ice-cream store. When the government levies the tax on sellers, the seller is required to place $0.50 in the bowl after the sale of each cone. When the government levies the tax on buyers, the buyer is required to place $0.50 in the bowl every time a cone is bought. Whether the $0.50 goes directly from the buyer’s pocket into the bowl, or indirectly from the buyer’s pocket into the seller’s hand and then into the bowl, does not matter. Once the market reaches its new equilibrium, buyers and sellers share the burden, regardless of how the tax is levied.

**CAN CONGRESS DISTRIBUTE THE BURDEN OF A PAYROLL TAX?**

If you have ever received a paycheck, you probably noticed that taxes were deducted from the amount you earned. One of these taxes is called FICA, an acronym for the Federal Insurance Contributions Act. The federal government uses the revenue from the FICA tax to pay for Social Security and Medicare, the income support and healthcare programs for the elderly. FICA is an example of a payroll tax, which is a tax on the wages that firms pay their workers. In 2008, the total FICA tax for the typical worker was 15.3 percent of earnings.

Who do you think bears the burden of this payroll tax—firms or workers? When Congress passed this legislation, it tried to mandate a division of the tax burden. According to the law, half of the tax is paid by firms, and half is paid by workers. That is, half of the tax is paid out of firms’ revenues, and half is deducted from workers’ paychecks. The amount that shows up as a deduction on your pay stub is the worker contribution.

Our analysis of tax incidence, however, shows that lawmakers cannot so easily dictate the distribution of a tax burden. To illustrate, we can analyze a payroll tax as merely a tax on a good, where the good is labor and the price is the wage. The key feature of the payroll tax is that it places a wedge between the wage that firms pay and the wage that workers receive. Figure 8 shows the outcome. When a payroll tax is enacted, the wage received by workers falls, and the wage paid by firms rises. In the end, workers and firms share the burden of the tax, much as the legislation requires. Yet this division of the tax burden between workers and firms has nothing to do with the legislated division: The division of the burden in Figure 8 is not necessarily fifty-fifty, and the same outcome would prevail if the law levied the entire tax on workers or if it levied the entire tax on firms.

This example shows that the most basic lesson of tax incidence is often overlooked in public debate. Lawmakers can decide whether a tax comes from the buyer’s pocket or from the seller’s, but they cannot legislate the true burden of a tax. Rather, tax incidence depends on the forces of supply and demand.  

●
Elasticity and Tax Incidence

When a good is taxed, buyers and sellers of the good share the burden of the tax. But how exactly is the tax burden divided? Only rarely will it be shared equally. To see how the burden is divided, consider the impact of taxation in the two markets in Figure 9. In both cases, the figure shows the initial demand curve, the initial supply curve, and a tax that drives a wedge between the amount paid by buyers and the amount received by sellers. (Not drawn in either panel of the figure is the new supply or demand curve. Which curve shifts depends on whether the tax is levied on buyers or sellers. As we have seen, this is irrelevant for the incidence of the tax.) The difference in the two panels is the relative elasticity of supply and demand.

Panel (a) of Figure 9 shows a tax in a market with very elastic supply and relatively inelastic demand. That is, sellers are very responsive to changes in the price of the good (so the supply curve is relatively flat), whereas buyers are not very responsive (so the demand curve is relatively steep). When a tax is imposed on a market with these elasticities, the price received by sellers does not fall much, so sellers bear only a small burden. By contrast, the price paid by buyers rises substantially, indicating that buyers bear most of the burden of the tax.

Panel (b) of Figure 9 shows a tax in a market with relatively inelastic supply and very elastic demand. In this case, sellers are not very responsive to changes in the price (so the supply curve is steeper), whereas buyers are very responsive (so the demand curve is flatter). The figure shows that when a tax is imposed, the price paid by buyers does not rise much, but the price received by sellers falls substantially. Thus, sellers bear most of the burden of the tax.

The two panels of Figure 9 show a general lesson about how the burden of a tax is divided: A tax burden falls more heavily on the side of the market that is less elastic. Why is this true? In essence, the elasticity measures the willingness of buyers or sellers to change their behavior in response to a change in the price of the good.
sellers to leave the market when conditions become unfavorable. A small elasticity of demand means that buyers do not have good alternatives to consuming this particular good. A small elasticity of supply means that sellers do not have good alternatives to producing this particular good. When the good is taxed, the side of the market with fewer good alternatives is less willing to leave the market and must, therefore, bear more of the burden of the tax.

We can apply this logic to the payroll tax discussed in the previous case study. Most labor economists believe that the supply of labor is much less elastic than the demand. This means that workers, rather than firms, bear most of the burden of
the payroll tax. In other words, the distribution of the tax burden is not at all close to the fifty-fifty split that lawmakers intended.

**WHO PAYS THE LUXURY TAX?**

In 1990, Congress adopted a new luxury tax on items such as yachts, private airplanes, furs, jewelry, and expensive cars. The goal of the tax was to raise revenue from those who could most easily afford to pay. Because only the rich could afford to buy such extravagances, taxing luxuries seemed a logical way of taxing the rich.

Yet, when the forces of supply and demand took over, the outcome was quite different from what Congress intended. Consider, for example, the market for yachts. The demand for yachts is quite elastic. A millionaire can easily not buy a yacht; she can use the money to buy a bigger house, take a European vacation, or leave a larger bequest to her heirs. By contrast, the supply of yachts is relatively inelastic, at least in the short run. Yacht factories are not easily converted to alternative uses, and workers who build yachts are not eager to change careers in response to changing market conditions.

Our analysis makes a clear prediction in this case. With elastic demand and inelastic supply, the burden of a tax falls largely on the suppliers. That is, a tax on yachts places a burden largely on the firms and workers who build yachts because they end up getting a significantly lower price for their product. The workers, however, are not wealthy. Thus, the burden of a luxury tax falls more on the middle class than on the rich.

The mistaken assumptions about the incidence of the luxury tax quickly became apparent after the tax went into effect. Suppliers of luxuries made their congressional representatives well aware of the economic hardship they experienced, and Congress repealed most of the luxury tax in 1993.

**QUICK QUIZ** In a supply-and-demand diagram, show how a tax on car buyers of $1,000 per car affects the quantity of cars sold and the price of cars. In another diagram, show how a tax on car sellers of $1,000 per car affects the quantity of cars sold and the price of cars. In both of your diagrams, show the change in the price paid by car buyers and the change in the price received by car sellers.

**CONCLUSION**

The economy is governed by two kinds of laws: the laws of supply and demand and the laws enacted by governments. In this chapter, we have begun to see how these laws interact. Price controls and taxes are common in various markets in the economy, and their effects are frequently debated in the press and among policymakers. Even a little bit of economic knowledge can go a long way toward understanding and evaluating these policies.

In subsequent chapters, we analyze many government policies in greater detail. We will examine the effects of taxation more fully, and we will consider a broader range of policies than we considered here. Yet the basic lessons of this chapter will not change: When analyzing government policies, supply and demand are the first and most useful tools of analysis.
SUMMARY

- A tax on a good places a wedge between the price paid by buyers and the price received by sellers. When the market moves to the new equilibrium, buyers pay more for the good and sellers receive less for it. In this sense, buyers and sellers share the tax burden. The incidence of a tax (that is, the division of the tax burden) does not depend on whether the tax is levied on buyers or sellers.

- The incidence of a tax depends on the price elasticities of supply and demand. Most of the burden falls on the side of the market that is less elastic because that side of the market can respond less easily to the tax by changing the quantity bought or sold.

- A price ceiling is a legal maximum on the price of a good or service. An example is rent control. If the price ceiling is below the equilibrium price, the quantity demanded exceeds the quantity supplied. Because of the resulting shortage, sellers must in some way ration the good or service among buyers.

- A price floor is a legal minimum on the price of a good or service. An example is the minimum wage. If the price floor is above the equilibrium price, the quantity supplied exceeds the quantity demanded. Because of the resulting surplus, buyers’ demands for the good or service must in some way be rationed among sellers.

- When the government levies a tax on a good, the equilibrium quantity of the good falls. That is, a tax on a market shrinks the size of the market.

KEY CONCEPTS

price ceiling, p. 114  price floor, p. 114  tax incidence, p. 124

QUESTIONS FOR REVIEW

1. Give an example of a price ceiling and an example of a price floor.
2. Which causes a shortage of a good—a price ceiling or a price floor? Justify your answer with a graph.
3. What mechanisms allocate resources when the price of a good is not allowed to bring supply and demand into equilibrium?
4. Explain why economists usually oppose controls on prices.
5. Suppose the government removes a tax on buyers of a good and levies a tax of the same size on sellers of the good. How does this change in tax policy affect the price that buyers pay sellers for this good, the amount buyers are out of pocket including the tax, the amount sellers receive net of the tax, and the quantity of the good sold?
6. How does a tax on a good affect the price paid by buyers, the price received by sellers, and the quantity sold?
7. What determines how the burden of a tax is divided between buyers and sellers? Why?
1. Lovers of classical music persuade Congress to impose a price ceiling of $40 per concert ticket. As a result of this policy, do more or fewer people attend classical music concerts?

2. The government has decided that the free-market price of cheese is too low.
   a. Suppose the government imposes a binding price floor in the cheese market. Draw a supply-and-demand diagram to show the effect of this policy on the price of cheese and the quantity of cheese sold. Is there a shortage or surplus of cheese?
   b. Farmers complain that the price floor has reduced their total revenue. Is this possible? Explain.
   c. In response to farmers’ complaints, the government agrees to purchase all the surplus cheese at the price floor. Compared to the basic price floor, who benefits from this new policy? Who loses?

3. A recent study found that the demand and supply schedules for Frisbees are as follows:

<table>
<thead>
<tr>
<th>Price per Frisbee</th>
<th>Quantity Demanded</th>
<th>Quantity Supplied</th>
</tr>
</thead>
<tbody>
<tr>
<td>$11</td>
<td>1 million</td>
<td>15 million</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

   a. What are the equilibrium price and quantity of Frisbees?
   b. Frisbee manufacturers persuade the government that Frisbee production improves scientists’ understanding of aerodynamics and thus is important for national security. A concerned Congress votes to impose a price floor $2 above the equilibrium price. What is the new market price? How many Frisbees are sold?
   c. Irate college students march on Washington and demand a reduction in the price of Frisbees. An even more concerned Congress votes to repeal the price floor and impose a price ceiling $1 below the former price floor. What is the new market price? How many Frisbees are sold?

4. Suppose the federal government requires beer drinkers to pay a $2 tax on each case of beer purchased. (In fact, both the federal and state governments impose beer taxes of some sort.)
   a. Draw a supply-and-demand diagram of the market for beer without the tax. Show the price paid by consumers, the price received by producers, and the quantity of beer sold. What is the difference between the price paid by consumers and the price received by producers?
   b. Now draw a supply-and-demand diagram for the beer market with the tax. Show the price paid by consumers, the price received by producers, and the quantity of beer sold. What is the difference between the price paid by consumers and the price received by producers? Has the quantity of beer sold increased or decreased?

5. A senator wants to raise tax revenue and make workers better off. A staff member proposes raising the payroll tax paid by firms and using part of the extra revenue to reduce the payroll tax paid by workers. Would this accomplish the senator’s goal? Explain.

6. If the government places a $500 tax on luxury cars, will the price paid by consumers rise by more than $500, less than $500, or exactly $500? Explain.

7. Congress and the president decide that the United States should reduce air pollution by reducing its use of gasoline. They impose a $0.50 tax for each gallon of gasoline sold.
   a. Should they impose this tax on producers or consumers? Explain carefully using a supply-and-demand diagram.
   b. If the demand for gasoline were more elastic, would this tax be more effective or less effective in reducing the quantity of gasoline consumed? Explain with both words and a diagram.
   c. Are consumers of gasoline helped or hurt by this tax? Why?
   d. Are workers in the oil industry helped or hurt by this tax? Why?

8. A case study in this chapter discusses the federal minimum-wage law.
a. Suppose the minimum wage is above the equilibrium wage in the market for unskilled labor. Using a supply-and-demand diagram of the market for unskilled labor, show the market wage, the number of workers who are employed, and the number of workers who are unemployed. Also show the total wage payments to unskilled workers.
b. Now suppose the secretary of labor proposes an increase in the minimum wage. What effect would this increase have on employment? Does the change in employment depend on the elasticity of demand, the elasticity of supply, both elasticities, or neither?
c. What effect would this increase in the minimum wage have on unemployment? Does the change in unemployment depend on the elasticity of demand, the elasticity of supply, both elasticities, or neither?
d. If the demand for unskilled labor were inelastic, would the proposed increase in the minimum wage raise or lower total wage payments to unskilled workers? Would your answer change if the demand for unskilled labor were elastic?

9. Consider the following policies, each of which is aimed at reducing violent crime by reducing the use of guns. Illustrate each of these proposed policies in a supply-and-demand diagram of the gun market.
   a. a tax on gun buyers
   b. a tax on gun sellers
   c. a price floor on guns
   d. a tax on ammunition

10. In 2007, Rod Blagojevich, the governor of Illinois, proposed a 3 percent payroll tax to finance some state health programs. The proposed legislation provided that the payroll tax “shall not be withheld from wages paid to employees or otherwise be collected from employees or reduce the compensation paid to employees.” What do you think was the intent of this language? Would the bill in fact have accomplished this objective?

11. The U.S. government administers two programs that affect the market for cigarettes. Media campaigns and labeling requirements are aimed at making the public aware of the dangers of cigarette smoking. At the same time, the Department of Agriculture maintains a price-support program for tobacco farmers, which raises the price of tobacco above the equilibrium price.
   a. How do these two programs affect cigarette consumption? Use a graph of the cigarette market in your answer.
   b. What is the combined effect of these two programs on the price of cigarettes?
   c. Cigarettes are also heavily taxed. What effect does this tax have on cigarette consumption?

12. At Fenway Park, home of the Boston Red Sox, seating is limited to 34,000. Hence, the number of tickets issued is fixed at that figure. (Assume that all seats are equally desirable and are sold at the same price.) Seeing a golden opportunity to raise revenue, the City of Boston levies a per ticket tax of $5 to be paid by the ticket buyer. Boston sports fans, a famously civic-minded lot, dutifully send in the $5 per ticket. Draw a well-labeled graph showing the impact of the tax. On whom does the tax burden fall—the team’s owners, the fans, or both? Why?

13. A subsidy is the opposite of a tax. With a $0.50 tax on the buyers of ice-cream cones, the government collects $0.50 for each cone purchased; with a $0.50 subsidy for the buyers of ice-cream cones, the government pays buyers $0.50 for each cone purchased.
   a. Show the effect of a $0.50 per cone subsidy on the demand curve for ice-cream cones, the effective price paid by consumers, the effective price received by sellers, and the quantity of cones sold.
   b. Do consumers gain or lose from this policy? Do producers gain or lose? Does the government gain or lose?

14. In the spring of 2008, Senators John McCain and Hillary Clinton (who were then running for President) proposed a temporary elimination of the federal gasoline tax, effective only during the summer of 2008, in order to help consumers deal with high gasoline prices.
   a. During the summer, when gasoline demand is high because of vacation driving, gasoline refiners are operating near full capacity. What does this fact suggest about the price elasticity of supply?
   b. In light of your answer to (a), who do you predict would benefit from the temporary gas tax holiday?
This page intentionally left blank
PART III
Markets and Welfare
This page intentionally left blank
Consumers, Producers, and the Efficiency of Markets

When consumers go to grocery stores to buy their turkeys for Thanksgiving dinner, they may be disappointed that the price of turkey is as high as it is. At the same time, when farmers bring to market the turkeys they have raised, they wish the price of turkey were even higher. These views are not surprising: Buyers always want to pay less, and sellers always want to be paid more. But is there a “right price” for turkey from the standpoint of society as a whole?

In previous chapters, we saw how, in market economies, the forces of supply and demand determine the prices of goods and services and the quantities sold. So far, however, we have described the way markets allocate scarce resources without directly addressing the question of whether these market allocations are desirable. In other words, our analysis has been positive (what is) rather than normative (what should be). We know that the price of turkey adjusts to ensure that the quantity of turkey supplied equals the quantity of turkey demanded. But at this equilibrium, is the quantity of turkey produced and consumed too small, too large, or just right?

In this chapter, we take up the topic of welfare economics, the study of how the allocation of resources affects economic well-being. We begin by examining welfare economics the study of how the allocation of resources affects economic well-being
the benefits that buyers and sellers receive from taking part in a market. We then examine how society can make these benefits as large as possible. This analysis leads to a profound conclusion: The equilibrium of supply and demand in a market maximizes the total benefits received by buyers and sellers.

As you may recall from Chapter 1, one of the Ten Principles of Economics is that markets are usually a good way to organize economic activity. The study of welfare economics explains this principle more fully. It also answers our question about the right price of turkey: The price that balances the supply and demand for turkey is, in a particular sense, the best one because it maximizes the total welfare of turkey consumers and turkey producers. No consumer or producer of turkeys aims to achieve this goal, but their joint action directed by market prices moves them toward a welfare-maximizing outcome, as if led by an invisible hand.

CONSUMER SURPLUS

We begin our study of welfare economics by looking at the benefits buyers receive from participating in a market.

Willingness to Pay

Imagine that you own a mint-condition recording of Elvis Presley’s first album. Because you are not an Elvis Presley fan, you decide to sell it. One way to do so is to hold an auction.

Four Elvis fans show up for your auction: John, Paul, George, and Ringo. Each of them would like to own the album, but there is a limit to the amount that each is willing to pay for it. Table 1 shows the maximum price that each of the four possible buyers would pay. Each buyer’s maximum is called his willingness to pay, and it measures how much that buyer values the good. Each buyer would be eager to buy the album at a price less than his willingness to pay, and he would refuse to buy the album at a price greater than his willingness to pay. At a price equal to his willingness to pay, the buyer would be indifferent about buying the good: If the price is exactly the same as the value he places on the album, he would be equally happy buying it or keeping his money.

To sell your album, you begin the bidding at a low price, say, $10. Because all four buyers are willing to pay much more, the price rises quickly. The bidding stops when John bids $80 (or slightly more). At this point, Paul, George, and Ringo have dropped out of the bidding because they are unwilling to bid any more than $80. John pays you $80 and gets the album. Note that the album has gone to the buyer who values the album most highly.

### Table 1

<table>
<thead>
<tr>
<th>Buyer</th>
<th>Willingness to Pay</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>$100</td>
</tr>
<tr>
<td>Paul</td>
<td>80</td>
</tr>
<tr>
<td>George</td>
<td>70</td>
</tr>
<tr>
<td>Ringo</td>
<td>50</td>
</tr>
</tbody>
</table>
What benefit does John receive from buying the Elvis Presley album? In a sense, John has found a real bargain: He is willing to pay $100 for the album but pays only $80 for it. We say that John receives consumer surplus of $20. Consumer surplus is the amount a buyer is willing to pay for a good minus the amount the buyer actually pays for it.

Consumer surplus measures the benefit buyers receive from participating in a market. In this example, John receives a $20 benefit from participating in the auction because he pays only $80 for a good he values at $100. Paul, George, and Ringo get no consumer surplus from participating in the auction because they left without the album and without paying anything.

Now consider a somewhat different example. Suppose that you had two identical Elvis Presley albums to sell. Again, you auction them off to the four possible buyers. To keep things simple, we assume that both albums are to be sold for the same price and that no buyer is interested in buying more than one album. Therefore, the price rises until two buyers are left.

In this case, the bidding stops when John and Paul bid $70 (or slightly higher). At this price, John and Paul are each happy to buy an album, and George and Ringo are not willing to bid any higher. John and Paul each receive consumer surplus equal to his willingness to pay minus the price. John’s consumer surplus is $30, and Paul’s is $10. John’s consumer surplus is higher now than in the previous example because he gets the same album but pays less for it. The total consumer surplus in the market is $40.

**USING THE DEMAND CURVE TO MEASURE CONSUMER SURPLUS**

Consumer surplus is closely related to the demand curve for a product. To see how they are related, let’s continue our example and consider the demand curve for this rare Elvis Presley album.

We begin by using the willingness to pay of the four possible buyers to find the demand schedule for the album. The table in Figure 1 shows the demand schedule that corresponds to Table 1. If the price is above $100, the quantity demanded in the market is 0 because no buyer is willing to pay that much. If the price is between $80 and $100, the quantity demanded is 1 because only John is willing to pay such a high price. If the price is between $70 and $80, the quantity demanded is 2 because both John and Paul are willing to pay the price. We can continue this analysis for other prices as well. In this way, the demand schedule is derived from the willingness to pay of the four possible buyers.

The graph in Figure 1 shows the demand curve that corresponds to this demand schedule. Note the relationship between the height of the demand curve and the buyers’ willingness to pay. At any quantity, the price given by the demand curve shows the willingness to pay of the marginal buyer, the buyer who would leave the market first if the price were any higher. At a quantity of 4 albums, for instance, the demand curve has a height of $50, the price that Ringo (the marginal buyer) is willing to pay for an album. At a quantity of 3 albums, the demand curve has a height of $70, the price that George (who is now the marginal buyer) is willing to pay.

Because the demand curve reflects buyers’ willingness to pay, we can also use it to measure consumer surplus. Figure 2 uses the demand curve to compute consumer surplus in our two examples. In panel (a), the price is $80 (or slightly
above), and the quantity demanded is 1. Note that the area above the price and below the demand curve equals $20. This amount is exactly the consumer surplus we computed earlier when only 1 album is sold.

Panel (b) of Figure 2 shows consumer surplus when the price is $70 (or slightly above). In this case, the area above the price and below the demand curve equals the total area of the two rectangles: John’s consumer surplus at this price is $30 and Paul’s is $10. This area equals a total of $40. Once again, this amount is the consumer surplus we computed earlier.

The lesson from this example holds for all demand curves: The area below the demand curve and above the price measures the consumer surplus in a market. This is true because the height of the demand curve measures the value buyers place on the good, as measured by their willingness to pay for it. The difference between this willingness to pay and the market price is each buyer’s consumer surplus. Thus, the total area below the demand curve and above the price is the sum of the consumer surplus of all buyers in the market for a good or service.

**How a Lower Price Raises Consumer Surplus**

Because buyers always want to pay less for the goods they buy, a lower price makes buyers of a good better off. But how much does buyers’ well-being rise in response to a lower price? We can use the concept of consumer surplus to answer this question precisely.

Figure 3 shows a typical demand curve. You may notice that this curve gradually slopes downward instead of taking discrete steps as in the previous two
figures. In a market with many buyers, the resulting steps from each buyer dropping out are so small that they form, in essence, a smooth curve. Although this curve has a different shape, the ideas we have just developed still apply: Consumer surplus is the area above the price and below the demand curve. In panel (a), consumer surplus at a price of \( P_1 \) is the area of triangle ABC.

Now suppose that the price falls from \( P_1 \) to \( P_2 \), as shown in panel (b). The consumer surplus now equals area ADF. The increase in consumer surplus attributable to the lower price is the area BCFD.

This increase in consumer surplus is composed of two parts. First, those buyers who were already buying \( Q_1 \) of the good at the higher price \( P_1 \) are better off because they now pay less. The increase in consumer surplus of existing buyers is the reduction in the amount they pay; it equals the area of the rectangle BCED. Second, some new buyers enter the market because they are willing to buy the good at the lower price. As a result, the quantity demanded in the market increases from \( Q_1 \) to \( Q_2 \). The consumer surplus these newcomers receive is the area of the triangle CEF.

**What Does Consumer Surplus Measure?**

Our goal in developing the concept of consumer surplus is to make judgments about the desirability of market outcomes. Now that you have seen what consumer surplus is, let’s consider whether it is a good measure of economic well-being.
Imagine that you are a policymaker trying to design a good economic system. Would you care about the amount of consumer surplus? Consumer surplus, the amount that buyers are willing to pay for a good minus the amount they actually pay for it, measures the benefit that buyers receive from a good as the buyers themselves perceive it. Thus, consumer surplus is a good measure of economic well-being if policymakers want to respect the preferences of buyers.

In some circumstances, policymakers might choose not to care about consumer surplus because they do not respect the preferences that drive buyer behavior. For example, drug addicts are willing to pay a high price for heroin. Yet we would not say that addicts get a large benefit from being able to buy heroin at a low price (even though addicts might say they do). From the standpoint of society, willingness to pay in this instance is not a good measure of the buyers’ benefit, and consumer surplus is not a good measure of economic well-being, because addicts are not looking after their own best interests.

In most markets, however, consumer surplus does reflect economic well-being. Economists normally assume that buyers are rational when they make decisions. Rational people do the best they can to achieve their objectives, given their opportunities. Economists also normally assume that people’s preferences should be respected. In this case, consumers are the best judges of how much benefit they receive from the goods they buy.

**Quick Quiz** Draw a demand curve for turkey. In your diagram, show a price of turkey and the consumer surplus at that price. Explain in words what this consumer surplus measures.
PRODUCER SURPLUS

We now turn to the other side of the market and consider the benefits sellers receive from participating in a market. As you will see, our analysis of sellers’ welfare is similar to our analysis of buyers’ welfare.

COST AND THE WILLINGNESS TO SELL

Imagine now that you are a homeowner and you want to get your house painted. You turn to four sellers of painting services: Mary, Frida, Georgia, and Grandma. Each painter is willing to do the work for you if the price is right. You decide to take bids from the four painters and auction off the job to the painter who will do the work for the lowest price.

Each painter is willing to take the job if the price she would receive exceeds her cost of doing the work. Here the term cost should be interpreted as the painters’ opportunity cost: It includes the painters’ out-of-pocket expenses (for paint, brushes, and so on) as well as the value that the painters place on their own time. Table 2 shows each painter’s cost. Because a painter’s cost is the lowest price she would accept for her work, cost is a measure of her willingness to sell her services. Each painter would be eager to sell her services at a price greater than her cost, and she would refuse to sell her services at a price less than her cost. At a price exactly equal to her cost, she would be indifferent about selling her services: She would be equally happy getting the job or using her time and energy for another purpose.

When you take bids from the painters, the price might start high, but it quickly falls as the painters compete for the job. Once Grandma has bid $600 (or slightly less), she is the sole remaining bidder. Grandma is happy to do the job for this price because her cost is only $500. Mary, Frida, and Georgia are unwilling to do the job for less than $600. Note that the job goes to the painter who can do the work at the lowest cost.

What benefit does Grandma receive from getting the job? Because she is willing to do the work for $500 but gets $600 for doing it, we say that she receives producer surplus of $100. Producer surplus is the amount a seller is paid minus the cost of production. Producer surplus measures the benefit sellers receive from participating in a market.

Now consider a somewhat different example. Suppose that you have two houses that need painting. Again, you auction off the jobs to the four painters. To keep things simple, let’s assume that no painter is able to paint both houses and that you will pay the same amount to paint each house. Therefore, the price falls until two painters are left.

<table>
<thead>
<tr>
<th>Seller</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary</td>
<td>$900</td>
</tr>
<tr>
<td>Frida</td>
<td>800</td>
</tr>
<tr>
<td>Georgia</td>
<td>600</td>
</tr>
<tr>
<td>Grandma</td>
<td>500</td>
</tr>
</tbody>
</table>
In this case, the bidding stops when Georgia and Grandma each offer to do the job for a price of $800 (or slightly less). Georgia and Grandma are willing to do the work at this price, while Mary and Frida are not willing to bid a lower price. At a price of $800, Grandma receives producer surplus of $300, and Georgia receives producer surplus of $200. The total producer surplus in the market is $500.

**Using the Supply Curve to Measure Producer Surplus**

Just as consumer surplus is closely related to the demand curve, producer surplus is closely related to the supply curve. To see how, let’s continue our example.

We begin by using the costs of the four painters to find the supply schedule for painting services. The table in Figure 4 shows the supply schedule that corresponds to the costs in Table 2. If the price is below $500, none of the four painters is willing to do the job, so the quantity supplied is zero. If the price is between $500 and $600, only Grandma is willing to do the job, so the quantity supplied is 1. If the price is between $600 and $800, Grandma and Georgia are willing to do the job, so the quantity supplied is 2, and so on. Thus, the supply schedule is derived from the costs of the four painters.

The graph in Figure 4 shows the supply curve that corresponds to this supply schedule. Note that the height of the supply curve is related to the sellers’ costs. At any quantity, the price given by the supply curve shows the cost of the marginal seller, the seller who would leave the market first if the price were any lower. At a quantity of 4 houses, for instance, the supply curve has a height of $900, the cost that Mary (the marginal seller) incurs to provide her painting services. At a quan-
tity of 3 houses, the supply curve has a height of $800, the cost that Frida (who is now the marginal seller) incurs.

Because the supply curve reflects sellers’ costs, we can use it to measure producer surplus. Figure 5 uses the supply curve to compute producer surplus in our two examples. In panel (a), we assume that the price is $600. In this case, the quantity supplied is 1. Note that the area below the price and above the supply curve equals $100. This amount is exactly the producer surplus we computed earlier for Grandma.

Panel (b) of Figure 5 shows producer surplus at a price of $800. In this case, the area below the price and above the supply curve equals the total area of the two rectangles. This area equals $500, the producer surplus we computed earlier for Georgia and Grandma when two houses needed painting.

The lesson from this example applies to all supply curves: The area below the price and above the supply curve measures the producer surplus in a market. The logic is straightforward: The height of the supply curve measures sellers’ costs, and the difference between the price and the cost of production is each seller’s producer surplus. Thus, the total area is the sum of the producer surplus of all sellers.

**How a Higher Price Raises Producer Surplus**

You will not be surprised to hear that sellers always want to receive a higher price for the goods they sell. But how much does sellers’ well-being rise in response to

---

**Figure 5**

Measuring Producer Surplus with the Supply Curve

In panel (a), the price of the good is $600, and the producer surplus is $100. In panel (b), the price of the good is $800, and the producer surplus is $500.
a higher price? The concept of producer surplus offers a precise answer to this question.

Figure 6 shows a typical upward-sloping supply curve that would arise in a market with many sellers. Although this supply curve differs in shape from the previous figure, we measure producer surplus in the same way: Producer surplus is the area below the price and above the supply curve. In panel (a), the price is $P_1$, and producer surplus is the area of triangle ABC.

Panel (b) shows what happens when the price rises from $P_1$ to $P_2$. Producer surplus now equals area ADF. This increase in producer surplus has two parts. First, those sellers who were already selling $Q_1$ of the good at the lower price $P_1$ are better off because they now get more for what they sell. The increase in producer surplus for existing sellers equals the area of the rectangle BCED. Second, some new sellers enter the market because they are willing to produce the good at the higher price, resulting in an increase in the quantity supplied from $Q_1$ to $Q_2$. The producer surplus of these newcomers is the area of the triangle CEF.

As this analysis shows, we use producer surplus to measure the well-being of sellers in much the same way as we use consumer surplus to measure the well-being of buyers. Because these two measures of economic welfare are so similar, it is natural to use them together. And indeed, that is exactly what we do in the next section.

**Quick Quiz** Draw a supply curve for turkey. In your diagram, show a price of turkey and the producer surplus at that price. Explain in words what this producer surplus measures.
Consumer surplus and producer surplus are the basic tools that economists use to study the welfare of buyers and sellers in a market. These tools can help us address a fundamental economic question: Is the allocation of resources determined by free markets desirable?

**The Benevolent Social Planner**

To evaluate market outcomes, we introduce into our analysis a new, hypothetical character called the benevolent social planner. The benevolent social planner is an all-knowing, all-powerful, well-intentioned dictator. The planner wants to maximize the economic well-being of everyone in society. What should this planner do? Should he just leave buyers and sellers at the equilibrium that they reach naturally on their own? Or can he increase economic well-being by altering the market outcome in some way?

To answer this question, the planner must first decide how to measure the economic well-being of a society. One possible measure is the sum of consumer and producer surplus, which we call total surplus. Consumer surplus is the benefit that buyers receive from participating in a market, and producer surplus is the benefit that sellers receive. It is therefore natural to use total surplus as a measure of society’s economic well-being.

To better understand this measure of economic well-being, recall how we measure consumer and producer surplus. We define consumer surplus as

\[
\text{Consumer surplus} = \text{Value to buyers} - \text{Amount paid by buyers}.
\]

Similarly, we define producer surplus as

\[
\text{Producer surplus} = \text{Amount received by sellers} - \text{Cost to sellers}.
\]

When we add consumer and producer surplus together, we obtain

\[
\text{Total surplus} = (\text{Value to buyers} - \text{Amount paid by buyers}) \quad + \quad (\text{Amount received by sellers} - \text{Cost to sellers}).
\]

The amount paid by buyers equals the amount received by sellers, so the middle two terms in this expression cancel each other. As a result, we can write total surplus as

\[
\text{Total surplus} = \text{Value to buyers} - \text{Cost to sellers}.
\]

Total surplus in a market is the total value to buyers of the goods, as measured by their willingness to pay, minus the total cost to sellers of providing those goods.

If an allocation of resources maximizes total surplus, we say that the allocation exhibits **efficiency**. If an allocation is not efficient, then some of the potential gains from trade among buyers and sellers are not being realized. For example, an allocation is inefficient if a good is not being produced by the sellers with lowest cost. In this case, moving production from a high-cost producer to a low-cost producer will lower the total cost to sellers and raise total surplus. Similarly, an
allocation is inefficient if a good is not being consumed by the buyers who value it most highly. In this case, moving consumption of the good from a buyer with a low valuation to a buyer with a high valuation will raise total surplus.

In addition to efficiency, the social planner might also care about equality—that is, whether the various buyers and sellers in the market have a similar level of economic well-being. In essence, the gains from trade in a market are like a pie to be shared among the market participants. The question of efficiency concerns whether the pie is as big as possible. The question of equality concerns how the pie is sliced and how the portions are distributed among members of society. In this chapter, we concentrate on efficiency as the social planner’s goal. Keep in mind, however, that real policymakers often care about equality as well.

Evaluating the Market Equilibrium

Figure 7 shows consumer and producer surplus when a market reaches the equilibrium of supply and demand. Recall that consumer surplus equals the area above the price and under the demand curve and producer surplus equals the area below the price and above the supply curve. Thus, the total area between the supply and demand curves up to the point of equilibrium represents the total surplus in this market.

Is this equilibrium allocation of resources efficient? That is, does it maximize total surplus? To answer this question, recall that when a market is in equilibrium, the price determines which buyers and sellers participate in the market. Those buyers who value the good more than the price (represented by the segment AE on the demand curve) choose to buy the good; buyers who value it less than the price (represented by the segment EB) do not. Similarly, those sellers whose

---

**Figure 7:** Consumer and Producer Surplus in the Market Equilibrium

Total surplus—the sum of consumer and producer surplus—is the area between the supply and demand curves up to the equilibrium quantity.
costs are less than the price (represented by the segment CE on the supply curve) choose to produce and sell the good; sellers whose costs are greater than the price (represented by the segment ED) do not.

These observations lead to two insights about market outcomes:

1. Free markets allocate the supply of goods to the buyers who value them most highly, as measured by their willingness to pay.
2. Free markets allocate the demand for goods to the sellers who can produce them at the least cost.

Thus, given the quantity produced and sold in a market equilibrium, the social planner cannot increase economic well-being by changing the allocation of consumption among buyers or the allocation of production among sellers.

But can the social planner raise total economic well-being by increasing or decreasing the quantity of the good? The answer is no, as stated in this third insight about market outcomes:

3. Free markets produce the quantity of goods that maximizes the sum of consumer and producer surplus.

Figure 8 illustrates why this is true. To interpret this figure, keep in mind that the demand curve reflects the value to buyers and the supply curve reflects the cost to sellers. At any quantity below the equilibrium level, such as $Q_1$, the value to the marginal buyer exceeds the cost to the marginal seller. As a result, increasing the quantity produced and consumed raises total surplus. This continues to be true until the quantity reaches the equilibrium level. Similarly, at any quantity beyond the equilibrium level, such as $Q_2$, the value to the marginal buyer is less than the...
cost to the marginal seller. In this case, decreasing the quantity raises total surplus, and this continues to be true until quantity falls to the equilibrium level. To maximize total surplus, the social planner would choose the quantity where the supply and demand curves intersect.

Together, these three insights tell us that the market outcome makes the sum of consumer and producer surplus as large as it can be. In other words, the equilibrium outcome is an efficient allocation of resources. The benevolent social planner can, therefore, leave the market outcome just as he finds it. This policy of leaving well enough alone goes by the French expression *laissez faire,* which literally translates to “allow them to do.”

Society is lucky that the planner doesn’t need to intervene. Although it has been a useful exercise imagining what an all-knowing, all-powerful, well-intentioned dictator would do, let’s face it: Such characters are hard to come by. Dictators are rarely benevolent, and even if we found someone so virtuous, he would lack crucial information.

Suppose our social planner tried to choose an efficient allocation of resources on his own, instead of relying on market forces. To do so, he would need to know the value of a particular good to every potential consumer in the market and the cost of every potential producer. And he would need this information not only for this market but for every one of the many thousands of markets in the economy. The task is practically impossible, which explains why centrally planned economies never work very well.

The planner’s job becomes easy, however, once he takes on a partner: Adam Smith’s invisible hand of the marketplace. The invisible hand takes all the information about buyers and sellers into account and guides everyone in the market to the best outcome as judged by the standard of economic efficiency. It is, truly, a remarkable feat. That is why economists so often advocate free markets as the best way to organize economic activity.

**CASE STUDY**

**SHOULD THERE BE A MARKET IN ORGANS?**

On April 12, 2001, the front page of the *Boston Globe* ran the headline “How a Mother’s Love Helped Save Two Lives.” The newspaper told the story of Susan Stephens, a woman whose son needed a kidney transplant. When the doctor learned that the mother’s kidney was not compatible, he proposed a novel solution: If Stephens donated one of her kidneys to a stranger, her son would move to the top of the kidney waiting list. The mother accepted the deal, and soon two patients had the transplant they were waiting for.

The ingenuity of the doctor’s proposal and the nobility of the mother’s act cannot be doubted. But the story raises some intriguing questions. If the mother could trade a kidney for a kidney, would the hospital allow her to trade a kidney for an expensive, experimental cancer treatment that she could not otherwise afford? Should she be allowed to exchange her kidney for free tuition for her son at the hospital’s medical school? Should she be able to sell her kidney so she can use the cash to trade in her old Chevy for a new Lexus?

As a matter of public policy, our society makes it illegal for people to sell their organs. In essence, in the market for organs, the government has imposed a price ceiling of zero. The result, as with any binding price ceiling, is a shortage of the good. The deal in the Stephens case did not fall under this prohibition because no
Ticket Scalping

To allocate resources efficiently, an economy must get goods—including tickets to the Red Sox—to the consumers who value them most highly.

Like It or Not, Scalping Is a Force in the Free Market

By Charles Stein

Chip Case devotes a class each year to the reselling of sports tickets. He has a section in his economics textbook on the same subject.

But for Case, an economics professor at Wellesley College, the sale and scalping of sports tickets is more than an interesting theoretical pursuit. Like Margaret Mead, he has done plenty of firsthand research in the jungle, and he has the stories to prove it.

In 1984, Case waited in line for two nights on Causeway Street to get $11 tickets to one of the classic Celtics-Lakers championship series. The night before the climactic seventh game, he was in the shower when his daughter called out to him: “Dad, there’s a guy on the phone who wants to buy your Celtics tickets.” Case said he wasn’t selling. “But Dad,” his daughter added, “he’s willing to pay at least $1,000 apiece for them.”

Case was selling. An hour later, a limo arrived at the house to pick up two tickets—one that belonged to Case and one to a friend of his. The driver left behind $3,000.

To Case and other economists, tickets are a textbook case of the free market in action. When supply is limited and demand is not, prices rise and the people willing to pay more will eventually get their hands on the tickets. “As long as people can communicate, there will be trades,” said Case.

In the age of the Internet, buyers and sellers can link up online, through eBay or the sites devoted solely to ticket sales. But even in the pre-Internet era, the process worked, albeit more slowly. In 1984, the man who bought Case’s tickets was a rich New Yorker whose son attended a Boston private school. The man called a friend at the school, who called someone else, who eventually called Case. Where there is a will, there is a way.

Trading happens no matter how hard teams try to suppress it. The National Football League gives some of its Super Bowl tickets to its teams, and prohibits them from reselling. Yet many of those same tickets wind up back on the secondary market. Last season the league caught Minnesota Vikings head coach Mike Tice selling his tickets to a California ticket agency. “I regret it,” Tice told Sports Illustrated afterward. Or at least he regretted getting caught.

Like any good market, the one for tickets is remarkably sensitive to information. Case has a story about that, too. He was in Kenmore Square just before game four of last year’s playoff series between the Yankees and Red Sox. The Red Sox had dropped the first three games and there was no joy in Mudville. Scalpers were unloading tickets for the fourth game for only slightly more than face value. Tickets for a possible fifth game were going for even less.

But the Red Sox rallied to win game four in extra innings. By 2 that morning, said Case, top tickets for game five were already selling for more than $1,000 online. A bear market had become a bull market instantaneously.

As defenders of the free market, economists generally see nothing wrong with scalping. “Consenting adults should be able to make economic trades when they think it is to their mutual advantage,” said Greg Mankiw, a Harvard economics professor who recently stepped down as chairman of President Bush’s Council of Economic Advisers. Mankiw has a section about scalping in his own textbook.

Teams could eliminate scalping altogether by holding their own online auctions for desirable tickets. Case doesn’t expect that to happen. “People would burn down Fenway Park if the Red Sox charged $2,000 for a ticket,” he said. The team would be accused of price gouging. Yet if you went online last week, you could find front-row Green Monster seats for the July 15 game against the Yankees selling for more than $2,000. Go figure.

Case will be at Fenway Park this Friday. He is taking his father-in-law to the game. He paid a small fortune for the tickets online. But he isn’t complaining. It’s the free market at work.

Source: Boston Globe, May 1, 2005.
Many economists believe that there would be large benefits to allowing a free market in organs. People are born with two kidneys, but they usually need only one. Meanwhile, a few people suffer from illnesses that leave them without any working kidney. Despite the obvious gains from trade, the current situation is dire: The typical patient has to wait several years for a kidney transplant, and every year thousands of people die because a kidney cannot be found. If those needing a kidney were allowed to buy one from those who have two, the price would rise to balance supply and demand. Sellers would be better off with the extra cash in their pockets. Buyers would be better off with the organ they need to save their lives. The shortage of kidneys would disappear.

Such a market would lead to an efficient allocation of resources, but critics of this plan worry about fairness. A market for organs, they argue, would benefit the rich at the expense of the poor because organs would then be allocated to those most willing and able to pay. But you can also question the fairness of the current system. Now, most of us walk around with an extra organ that we don’t really need, while some of our fellow citizens are dying to get one. Is that fair?

**Quick Quiz** Draw the supply and demand for turkey. In the equilibrium, show producer and consumer surplus. Explain why producing more turkeys would lower total surplus.

---

**CONCLUSION: MARKET EFFICIENCY AND MARKET FAILURE**

This chapter introduced the basic tools of welfare economics—consumer and producer surplus—and used them to evaluate the efficiency of free markets. We showed that the forces of supply and demand allocate resources efficiently. That is, even though each buyer and seller in a market is concerned only about his or her own welfare, they are together led by an invisible hand to an equilibrium that maximizes the total benefits to buyers and sellers.

A word of warning is in order. To conclude that markets are efficient, we made several assumptions about how markets work. When these assumptions do not hold, our conclusion that the market equilibrium is efficient may no longer be true. As we close this chapter, let’s consider briefly two of the most important of these assumptions.

First, our analysis assumed that markets are perfectly competitive. In the world, however, competition is sometimes far from perfect. In some markets, a single buyer or seller (or a small group of them) may be able to control market prices. This ability to influence prices is called market power. Market power can cause markets to be inefficient because it keeps the price and quantity away from the equilibrium of supply and demand.

Second, our analysis assumed that the outcome in a market matters only to the buyers and sellers in that market. Yet, in the world, the decisions of buyers and sellers sometimes affect people who are not participants in the market at all. Pollution is the classic example. The use of agricultural pesticides, for instance, affects not only the manufacturers who make them and the farmers who use them, but many others who breathe air or drink water that has been polluted with these pesticides. Such side effects, called externalities, cause welfare in a market to depend on more than just the value to the buyers and the cost to the sellers. Because buy-
The Miracle of the Market
An opinion columnist suggests that the next time you sit down for Thanksgiving dinner, you should give thanks not only for the turkey on your plate but also for the economic system in which you live.

Giving Thanks for the “Invisible Hand”
By Jeff Jacoby

Gratitude to the Almighty is the theme of Thanksgiving, and has been ever since the Pilgrims of Plymouth brought in their first good harvest. . . . Today, in millions of homes across the nation, God will be thanked for many gifts—for the feast on the table and the company of loved ones, for health and good fortune in the year gone by, for peace at home in a time of war, for the incalculable privilege of having been born—or having become—American.

But it probably won’t occur to too many of us to give thanks for the fact that the local supermarket had plenty of turkey for sale this week. Even the devout aren’t likely to thank God for airline schedules that made it possible for some of those loved ones to fly home for Thanksgiving. Or for the arrival of “Master and Commander” at the local movie theater in time for the holiday weekend. Or for that great cranberry-apple pie recipe in the food section of the newspaper.

Those things we take more or less for granted. It hardly takes a miracle to explain why grocery stores stock up on turkey before Thanksgiving, or why Hollywood releases big movies in time for big holidays. That’s what they do. Where is God in that?

And yet, isn’t there something wondrous—something almost inexplicable—in the way your Thanksgiving weekend is made possible by the skill and labor of vast numbers of total strangers?

To bring that turkey to the dining room table, for example, required the efforts of thousands of people—the poultry farmers who raised the birds, of course, but also the feed distributors who supplied their nourishment and the truckers who brought it to the farm, not to mention the architect who designed the hatchery, the workmen who built it, and the technicians who keep it running. The bird had to be slaughtered and defeathered and inspected and transported and unloaded and wrapped and priced and displayed. The people who accomplished those tasks were supported in turn by armies of other people accomplishing other tasks—from refining the gasoline that fueled the trucks to manufacturing the plastic in which the meat was packaged.

The activities of countless far-flung men and women over the course of many months had to be intricately choreographed and precisely timed, so that when you showed up to buy a fresh Thanksgiving turkey, there would be one—or more likely, a few dozen—waiting. The level of coordination that was required to pull it off is mind-boggling. But what is even more mind-boggling is this: No one coordinated it.

No turkey czar sat in a command post somewhere, consulting a master plan and issuing orders. No one rode herd on all those people, forcing them to cooperate for your benefit. And yet they did cooperate. When you arrived at the supermarket, your turkey was there. You didn’t have to do anything but show up to buy it. If that isn’t a miracle, what should we call it?

Adam Smith called it “the invisible hand”—the mysterious power that leads innumerable people, each working for his own gain, to promote ends that benefit many. Out of the seeming chaos of millions of uncoordinated private transactions emerges the spontaneous order of the market. Free human beings freely interact, and the result is an array of goods and services more immense than the human mind can comprehend. No dictator, no bureaucracy, no supercomputer plans it in advance. Indeed, the more an economy is planned, the more it is plagued by shortages, dislocation, and failure. . . .

The social order of freedom, like the wealth and the progress it makes possible, is an extraordinary gift from above. On this Thanksgiving Day and every day, may we be grateful.

ers and sellers do not consider these side effects when deciding how much to consume and produce, the equilibrium in a market can be inefficient from the standpoint of society as a whole.

Market power and externalities are examples of a general phenomenon called *market failure*—the inability of some unregulated markets to allocate resources efficiently. When markets fail, public policy can potentially remedy the problem and increase economic efficiency. Microeconomists devote much effort to studying when market failure is likely and what sorts of policies are best at correcting market failures. As you continue your study of economics, you will see that the tools of welfare economics developed here are readily adapted to that endeavor.

Despite the possibility of market failure, the invisible hand of the marketplace is extraordinarily important. In many markets, the assumptions we made in this chapter work well, and the conclusion of market efficiency applies directly. Moreover, we can use our analysis of welfare economics and market efficiency to shed light on the effects of various government policies. In the next two chapters, we apply the tools we have just developed to study two important policy issues—the welfare effects of taxation and of international trade.

**SUMMARY**

- Consumer surplus equals buyers’ willingness to pay for a good minus the amount they actually pay, and it measures the benefit buyers get from participating in a market. Consumer surplus can be computed by finding the area below the demand curve and above the price.

- Producer surplus equals the amount sellers receive for their goods minus their costs of production, and it measures the benefit sellers get from participating in a market. Producer surplus can be computed by finding the area below the price and above the supply curve.

- An allocation of resources that maximizes the sum of consumer and producer surplus is said to be efficient. Policymakers are often concerned with the efficiency, as well as the equality, of economic outcomes.

- The equilibrium of supply and demand maximizes the sum of consumer and producer surplus. That is, the invisible hand of the marketplace leads buyers and sellers to allocate resources efficiently.

- Markets do not allocate resources efficiently in the presence of market failures such as market power or externalities.
1. Explain how buyers’ willingness to pay, consumer surplus, and the demand curve are related.
2. Explain how sellers’ costs, producer surplus, and the supply curve are related.
3. In a supply-and-demand diagram, show producer and consumer surplus in the market equilibrium.
4. What is efficiency? Is it the only goal of economic policymakers?
5. What does the invisible hand do?
6. Name two types of market failure. Explain why each may cause market outcomes to be inefficient.

1. Melissa buys an iPod for $120 and gets consumer surplus of $80.
   a. What is her willingness to pay?
   b. If she had bought the iPod on sale for $90, what would her consumer surplus have been?
   c. If the price of an iPod were $250, what would her consumer surplus have been?
4. It is a hot day, and Bert is thirsty. Here is the value he places on a bottle of water:
   - Value of first bottle: $7
   - Value of second bottle: $5
   - Value of third bottle: $3
   - Value of fourth bottle: $1
   a. From this information, derive Bert’s demand schedule. Graph his demand curve for bottled water.
   b. If the price of a bottle of water is $4, how many bottles does Bert buy? How much consumer surplus does Bert get from his purchases? Show Bert’s consumer surplus in your graph.
   c. If the price falls to $2, how does quantity demanded change? How does Bert’s consumer surplus change? Show these changes in your graph.
5. Ernie owns a water pump. Because pumping large amounts of water is harder than pumping small amounts, the cost of producing a bottle of water rises as he pumps more. Here is the cost he incurs to produce each bottle of water:
   - Cost of first bottle: $1
   - Cost of second bottle: $3
   - Cost of third bottle: $5
   - Cost of fourth bottle: $7
   a. From this information, derive Ernie’s supply schedule. Graph his supply curve for bottled water.
b. If the price of a bottle of water is $4, how many bottles does Ernie produce and sell? How much producer surplus does Ernie get from these sales? Show Ernie’s producer surplus in your graph.

c. If the price rises to $6, how does quantity supplied change? How does Ernie’s producer surplus change? Show these changes in your graph.

6. Consider a market in which Bert from Problem 4 is the buyer and Ernie from Problem 5 is the seller.

a. Use Ernie’s supply schedule and Bert’s demand schedule to find the quantity supplied and quantity demanded at prices of $2, $4, and $6. Which of these prices brings supply and demand into equilibrium?

b. What are consumer surplus, producer surplus, and total surplus in this equilibrium?

c. If Ernie produced and Bert consumed one fewer bottle of water, what would happen to total surplus?

d. If Ernie produced and Bert consumed one additional bottle of water, what would happen to total surplus?

7. The cost of producing flat-screen TVs has fallen over the past several decades. Let’s consider some implications of this fact.

a. Draw a supply-and-demand diagram to show the effect of falling production costs on the price and quantity of flat-screen TVs sold.

b. In your diagram, show what happens to consumer surplus and producer surplus.

c. Suppose the supply of flat-screen TVs is very elastic. Who benefits most from falling production costs—consumers or producers of these TVs?

8. There are four consumers willing to pay the following amounts for haircuts:

<table>
<thead>
<tr>
<th>Consumer</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jerry</td>
<td>$7</td>
</tr>
<tr>
<td>Oprah</td>
<td>$2</td>
</tr>
<tr>
<td>Ellen</td>
<td>$8</td>
</tr>
<tr>
<td>Phil</td>
<td>$5</td>
</tr>
</tbody>
</table>

There are four haircutting businesses with the following costs:

<table>
<thead>
<tr>
<th>Firm</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$3</td>
</tr>
<tr>
<td>B</td>
<td>$6</td>
</tr>
<tr>
<td>C</td>
<td>$4</td>
</tr>
<tr>
<td>D</td>
<td>$2</td>
</tr>
</tbody>
</table>

Each firm has the capacity to produce only one haircut. For efficiency, how many haircuts should be given? Which businesses should cut hair and which consumers should have their hair cut? How large is the maximum possible total surplus?

9. Suppose a technological advance reduces the cost of making computers.

a. Draw a supply-and-demand diagram to show what happens to price, quantity, consumer surplus, and producer surplus in the market for computers.

b. Computers and adding machines are substitutes. Use a supply-and-demand diagram to show what happens to price, quantity, consumer surplus, and producer surplus in the market for adding machines. Should adding machine producers be happy or sad about the technological advance in computers?

c. Computers and software are complements. Draw a supply-and-demand diagram to show what happens to price, quantity, consumer surplus, and producer surplus in the market for software. Should software producers be happy or sad about the technological advance in computers?

d. Does this analysis help explain why software producer Bill Gates is one of the world’s richest men?

10. Consider how health insurance affects the quantity of healthcare services performed. Suppose that the typical medical procedure has a cost of $100, yet a person with health insurance pays only $20 out of pocket. Her insurance company pays the remaining $80. (The insurance company recoups the $80
through premiums, but the premium a person pays does not depend on how many procedures that person chooses to undertake.)

a. Draw the demand curve in the market for medical care. (In your diagram, the horizontal axis should represent the number of medical procedures.) Show the quantity of procedures demanded if each procedure has a price of $100.

b. On your diagram, show the quantity of procedures demanded if consumers pay only $20 per procedure. If the cost of each procedure to society is truly $100, and if individuals have health insurance as just described, will the number of procedures performed maximize total surplus? Explain.

c. Economists often blame the health insurance system for excessive use of medical care.

given your analysis, why might the use of care be viewed as “excessive”?

d. What sort of policies might prevent this excessive use?

11. The supply and demand for broccoli are described by the following equations:

Supply: \( Q^s = 4P - 80 \)
Demand: \( Q^d = 100 - 2P \).

\( Q \) is in bushels, and \( P \) is in dollars per bushel.

a. Graph the supply curve and the demand curve. What is the equilibrium price and quantity?

b. Calculate consumer surplus, producer surplus, and total surplus at the equilibrium.

c. If a dictator who hated broccoli were to ban the vegetable, who would bear the larger burden—the buyers or sellers of broccoli?
This page intentionally left blank
Application: The Costs of Taxation

Taxes are often a source of heated political debate. In 1776, the anger of the American colonists over British taxes sparked the American Revolution. More than two centuries later, the American political parties continue to debate the proper size and shape of the tax system. Yet no one would deny that some level of taxation is necessary. As Oliver Wendell Holmes Jr. once said, "Taxes are what we pay for civilized society."

Because taxation has such a major impact on the modern economy, we return to the topic several times throughout this book as we expand the set of tools we have at our disposal. We began our study of taxes in Chapter 6. There we saw how a tax on a good affects its price and the quantity sold and how the forces of supply and demand divide the burden of a tax between buyers and sellers. In this chapter, we extend this analysis and look at how taxes affect welfare, the economic well-being of participants in a market. In other words, we see how high the price of civilized society can be.

The effects of taxes on welfare might at first seem obvious. The government enacts taxes to raise revenue, and that revenue must come out of someone’s pocket. As we saw in Chapter 6, both buyers and sellers are worse off when a good is taxed: A tax raises the price buyers pay and lowers the price sellers receive. Yet to
understand more fully how taxes affect economic well-being, we must compare the reduced welfare of buyers and sellers to the amount of revenue the government raises. The tools of consumer and producer surplus allow us to make this comparison. The analysis will show that the cost of taxes to buyers and sellers exceeds the revenue raised by the government.

THE DEADWEIGHT LOSS OF TAXATION

We begin by recalling one of the surprising lessons from Chapter 6: The outcome is the same whether a tax on a good is levied on buyers or sellers of the good. When a tax is levied on buyers, the demand curve shifts downward by the size of the tax; when it is levied on sellers, the supply curve shifts upward by that amount. In either case, when the tax is enacted, the price paid by buyers rises, and the price received by sellers falls. In the end, the elasticities of supply and demand determine how the tax burden is distributed between producers and consumers. This distribution is the same regardless of how it is levied.

Figure 1 shows these effects. To simplify our discussion, this figure does not show a shift in either the supply or demand curve, although one curve must shift. Which curve shifts depends on whether the tax is levied on sellers (the supply curve shifts) or buyers (the demand curve shifts). In this chapter, we can keep the analysis general and simplify the graphs by not bothering to show the shift. The key result for our purposes here is that the tax places a wedge between the price buyers pay and the price sellers receive. Because of this tax wedge, the quantity sold falls below the level that would be sold without a tax. In other words, a tax on a good causes the size of the market for the good to shrink. These results should be familiar from Chapter 6.
**How a Tax Affects Market Participants**

Let’s use the tools of welfare economics to measure the gains and losses from a tax on a good. To do this, we must take into account how the tax affects buyers, sellers, and the government. The benefit received by buyers in a market is measured by consumer surplus—the amount buyers are willing to pay for the good minus the amount they actually pay for it. The benefit received by sellers in a market is measured by producer surplus—the amount sellers receive for the good minus their costs. These are precisely the measures of economic welfare we used in Chapter 7.

What about the third interested party, the government? If \( T \) is the size of the tax and \( Q \) is the quantity of the good sold, then the government gets total tax revenue of \( T \times Q \). It can use this tax revenue to provide services, such as roads, police, and public education, or to help the needy. Therefore, to analyze how taxes affect economic well-being, we use the government’s tax revenue to measure the public benefit from the tax. Keep in mind, however, that this benefit actually accrues not to government but to those on whom the revenue is spent.

Figure 2 shows that the government’s tax revenue is represented by the rectangle between the supply and demand curves. The height of this rectangle is the size of the tax, \( T \), and the width of the rectangle is the quantity of the good sold, \( Q \). Because a rectangle’s area is its height times its width, this rectangle’s area is \( T \times Q \), which equals the tax revenue.

**Welfare without a Tax** To see how a tax affects welfare, we begin by considering welfare before the government imposes a tax. Figure 3 shows the supply-and-demand diagram and marks the key areas with the letters A through F.

Without a tax, the equilibrium price and quantity are found at the intersection of the supply and demand curves. The price is \( P_1 \), and the quantity sold is \( Q_1 \).

**Tax Revenue** The tax revenue that the government collects equals \( T \times Q \), the size of the tax \( T \) times the quantity sold \( Q \). Thus, tax revenue equals the area of the rectangle between the supply and demand curves.
Because the demand curve reflects buyers’ willingness to pay, consumer surplus is the area between the demand curve and the price, $A + B + C$. Similarly, because the supply curve reflects sellers’ costs, producer surplus is the area between the supply curve and the price, $D + E + F$. In this case, because there is no tax, tax revenue equals zero.

Total surplus, the sum of consumer and producer surplus, equals the area $A + B + C + D + E + F$. In other words, as we saw in Chapter 7, total surplus is the area between the supply and demand curves up to the equilibrium quantity. The first column of the table in Figure 3 summarizes these conclusions.

**Welfare with a Tax** Now consider welfare after the tax is enacted. The price paid by buyers rises from $P_1$ to $P_B$, so consumer surplus now equals only area $A$ (the area below the demand curve and above the buyer’s price). The price received by sellers falls from $P_1$ to $P_S$, so producer surplus now equals only area $F$ (the area above the supply curve and below the seller’s price). The quantity sold falls from $Q_1$ to $Q_2$, and the government collects tax revenue equal to the area $B + D$. 
To compute total surplus with the tax, we add consumer surplus, producer surplus, and tax revenue. Thus, we find that total surplus is area $A + B + D + F$. The second column of the table summarizes these results.

**Changes in Welfare**  We can now see the effects of the tax by comparing welfare before and after the tax is enacted. The third column of the table in Figure 3 shows the changes. The tax causes consumer surplus to fall by the area $B + C$ and producer surplus to fall by the area $D + E$. Tax revenue rises by the area $B + D$. Not surprisingly, the tax makes buyers and sellers worse off and the government better off.

The change in total welfare includes the change in consumer surplus (which is negative), the change in producer surplus (which is also negative), and the change in tax revenue (which is positive). When we add these three pieces together, we find that total surplus in the market falls by the area $C + E$. Thus, the losses to buyers and sellers from a tax exceed the revenue raised by the government. The fall in total surplus that results when a tax (or some other policy) distorts a market outcome is called the **deadweight loss**. The area $C + E$ measures the size of the deadweight loss.

To understand why taxes impose deadweight losses, recall one of the **Ten Principles of Economics** in Chapter 1: People respond to incentives. In Chapter 7, we saw that free markets normally allocate scarce resources efficiently. That is, the equilibrium of supply and demand maximizes the total surplus of buyers and sellers in a market. When a tax raises the price to buyers and lowers the price to sellers, however, it gives buyers an incentive to consume less and sellers an incentive to produce less than they would in the absence of the tax. As buyers and sellers respond to these incentives, the size of the market shrinks below its optimum (as shown in the figure by the movement from $Q_1$ to $Q_2$). Thus, because taxes distort incentives, they cause markets to allocate resources inefficiently.

---

**Deadweight Losses and the Gains from Trade**

To gain some intuition for why taxes result in deadweight losses, consider an example. Imagine that Joe cleans Jane’s house each week for $100. The opportunity cost of Joe’s time is $80, and the value of a clean house to Jane is $120. Thus, Joe and Jane each receive a $20 benefit from their deal. The total surplus of $40 measures the gains from trade in this particular transaction.

Now suppose that the government levies a $50 tax on the providers of cleaning services. There is now no price that Jane can pay Joe that will leave both of them better off after paying the tax. The most Jane would be willing to pay is $120, but then Joe would be left with only $70 after paying the tax, which is less than his $80 opportunity cost. Conversely, for Joe to receive his opportunity cost of $80, Jane would need to pay $130, which is above the $120 value she places on a clean house. As a result, Jane and Joe cancel their arrangement. Joe goes without the income, and Jane lives in a dirtier house.

The tax has made Joe and Jane worse off by a total of $40 because they have each lost $20 of surplus. But note that the government collects no revenue from Joe and Jane because they decide to cancel their arrangement. The $40 is pure deadweight loss: It is a loss to buyers and sellers in a market that is not offset by an increase in government revenue. From this example, we can see the ultimate source of deadweight losses: **Taxes cause deadweight losses because they prevent buyers and sellers from realizing some of the gains from trade.**
The area of the triangle between the supply and demand curves (area C + E in Figure 3) measures these losses. This conclusion can be seen more easily in Figure 4 by recalling that the demand curve reflects the value of the good to consumers and that the supply curve reflects the costs of producers. When the tax raises the price to buyers to $P_B$ and lowers the price to sellers to $P_S$, the marginal buyers and sellers leave the market, so the quantity sold falls from $Q_1$ to $Q_2$. Yet as the figure shows, the value of the good to these buyers still exceeds the cost to these sellers. At every quantity between $Q_1$ and $Q_2$, the situation is the same as in our example with Joe and Jane. The gains from trade—the difference between buyers’ value and sellers’ cost—are less than the tax. As a result, these trades are not made once the tax is imposed. The deadweight loss is the surplus lost because the tax discourages these mutually advantageous trades.

**QUICK QUIZ** Draw the supply and demand curves for cookies. If the government imposes a tax on cookies, show what happens to the price paid by buyers, the price received by sellers, and the quantity sold. In your diagram, show the deadweight loss from the tax. Explain the meaning of the deadweight loss.

---

**THE DETERMINANTS OF THE DEADWEIGHT LOSS**

What determines whether the deadweight loss from a tax is large or small? The answer is the price elasticities of supply and demand, which measure how much the quantity supplied and quantity demanded respond to changes in the price.

Let’s consider first how the elasticity of supply affects the size of the deadweight loss. In the top two panels of Figure 5, the demand curve and the size of the tax are the same. The only difference in these figures is the elasticity of the
supply curve. In panel (a), the supply curve is relatively inelastic: Quantity supplied responds only slightly to changes in the price. In panel (b), the supply curve is relatively elastic: Quantity supplied responds substantially to changes in the price. Notice that the deadweight loss, the area of the triangle between the supply and demand curves, is larger when the supply curve is more elastic.

Similarly, the bottom two panels of Figure 5 show how the elasticity of demand affects the size of the deadweight loss. Here the supply curve and the size of the

FIGURE 5

Tax Distortions and Elasticities

In panels (a) and (b), the demand curve and the size of the tax are the same, but the price elasticity of supply is different. Notice that the more elastic the supply curve, the larger the deadweight loss of the tax. In panels (c) and (d), the supply curve and the size of the tax are the same, but the price elasticity of demand is different. Notice that the more elastic the demand curve, the larger the deadweight loss of the tax.
tax are held constant. In panel (c), the demand curve is relatively inelastic, and the deadweight loss is small. In panel (d), the demand curve is more elastic, and the deadweight loss from the tax is larger.

The lesson from this figure is easy to explain. A tax has a deadweight loss because it induces buyers and sellers to change their behavior. The tax raises the price paid by buyers, so they consume less. At the same time, the tax lowers the price received by sellers, so they produce less. Because of these changes in behavior, the size of the market shrinks below the optimum. The elasticities of supply and demand measure how much sellers and buyers respond to the changes in the price and, therefore, determine how much the tax distorts the market outcome. Hence, the greater the elasticities of supply and demand, the greater the deadweight loss of a tax.

THE DEADWEIGHT LOSS DEBATE

Supply, demand, elasticity, deadweight loss—all this economic theory is enough to make your head spin. But believe it or not, these ideas go to the heart of a profound political question: How big should the government be? The debate hinges on these concepts because the larger the deadweight loss of taxation, the larger the cost of any government program. If taxation entails large deadweight losses, then these losses are a strong argument for a leaner government that does less and taxes less. But if taxes impose small deadweight losses, then government programs are less costly than they otherwise might be.

So how big are the deadweight losses of taxation? Economists disagree on the answer to this question. To see the nature of this disagreement, consider the most important tax in the U.S. economy: the tax on labor. The Social Security tax, the Medicare tax, and, to a large extent, the federal income tax are labor taxes. Many state governments also tax labor earnings. A labor tax places a wedge between the wage that firms pay and the wage that workers receive. For a typical worker, if all forms of labor taxes are added together, the marginal tax rate on labor income—the tax on the last dollar of earnings—is about 40 percent.

Although the size of the labor tax is easy to determine, the deadweight loss of this tax is less straightforward. Economists disagree about whether this 40 percent labor tax has a small or a large deadweight loss. This disagreement arises because economists hold different views about the elasticity of labor supply.

Economists who argue that labor taxes do not greatly distort market outcomes believe that labor supply is fairly inelastic. Most people, they claim, would work full time regardless of the wage. If so, the labor supply curve is almost vertical, and a tax on labor has a small deadweight loss.

Economists who argue that labor taxes are highly distorting believe that labor supply is more elastic. While admitting that some groups of workers may supply their labor inelastically, these economists claim that many other groups respond more to incentives. Here are some examples:

- Many workers can adjust the number of hours they work—for instance, by working overtime. The higher the wage, the more hours they choose to work.
- Some families have second earners—often married women with children—with some discretion over whether to do unpaid work at home or paid work in the marketplace. When deciding whether to take a job, these second earn-
ers compare the benefits of being at home (including savings on the cost of child care) with the wages they could earn.

- Many of the elderly can choose when to retire, and their decisions are partly based on the wage. Once they are retired, the wage determines their incentive to work part time.
- Some people consider engaging in illegal economic activity, such as the drug trade, or working at jobs that pay “under the table” to evade taxes. Economists call this the underground economy. In deciding whether to work in the underground economy or at a legitimate job, these potential criminals compare what they can earn by breaking the law with the wage they can earn legally.

In each of these cases, the quantity of labor supplied responds to the wage (the price of labor). Thus, the decisions of these workers are distorted when their labor earnings are taxed. Labor taxes encourage workers to work fewer hours, second earners to stay at home, the elderly to retire early, and the unscrupulous to enter the underground economy.

These two views of labor taxation persist to this day. Indeed, whenever you see two political candidates debating whether the government should provide more services or reduce the tax burden, keep in mind that part of the disagreement may rest on different views about the elasticity of labor supply and the deadweight loss of taxation.

**QUICK QUIZ** The demand for beer is more elastic than the demand for milk. Would a tax on beer or a tax on milk have a larger deadweight loss? Why?

---

## DEADWEIGHT LOSS AND TAX REVENUE AS TAXES VARY

Taxes rarely stay the same for long periods of time. Policymakers in local, state, and federal governments are always considering raising one tax or lowering another. Here we consider what happens to the deadweight loss and tax revenue when the size of a tax changes.

Figure 6 shows the effects of a small, medium, and large tax, holding constant the market’s supply and demand curves. The deadweight loss—the reduction in total surplus that results when the tax reduces the size of a market below the optimum—equals the area of the triangle between the supply and demand curves. For the small tax in panel (a), the area of the deadweight loss triangle is quite small. But as the size of a tax rises in panels (b) and (c), the deadweight loss grows larger and larger.

Indeed, the deadweight loss of a tax rises even more rapidly than the size of the tax. This occurs because the deadweight loss is an area of a triangle, and the area of a triangle depends on the square of its size. If we double the size of a tax, for instance, the base and height of the triangle double, so the deadweight loss rises by a factor of 4. If we triple the size of a tax, the base and height triple, so the deadweight loss rises by a factor of 9.

The government’s tax revenue is the size of the tax times the amount of the good sold. As the first three panels of Figure 6 show, tax revenue equals the area
The deadweight loss is the reduction in total surplus due to the tax. Tax revenue is the amount of the tax times the amount of the good sold. In panel (a), a small tax has a small deadweight loss and raises a small amount of revenue. In panel (b), a somewhat larger tax has a larger deadweight loss and raises a larger amount of revenue. In panel (c), a very large tax has a very large deadweight loss, but because it has reduced the size of the market so much, the tax raises only a small amount of revenue. Panels (d) and (e) summarize these conclusions. Panel (d) shows that as the size of a tax grows larger, the deadweight loss grows larger. Panel (e) shows that tax revenue first rises and then falls. This relationship is sometimes called the Laffer curve.

The last two panels of Figure 6 summarize these results. In panel (d), we see that as the size of a tax increases, its deadweight loss quickly gets larger. By con-
Henry George and the Land Tax

Is there an ideal tax? Henry George, the 19th-century American economist and social philosopher, thought so. In his 1879 book *Progress and Poverty*, George argued that the government should raise all its revenue from a tax on land. This “single tax” was, he claimed, both equitable and efficient. George’s ideas won him a large political following, and in 1886, he lost a close race for mayor of New York City (although he finished well ahead of Republican candidate Theodore Roosevelt).

George’s proposal to tax land was motivated largely by a concern over the distribution of economic well-being. He deplored the “shocking contrast between monstrous wealth and debasing want” and thought landowners benefited more than they should from the rapid growth in the overall economy.

George’s arguments for the land tax can be understood using the tools of modern economics. Consider first supply and demand in the market for renting land. As immigration causes the population to rise and technological progress causes incomes to grow, the demand for land rises over time. Yet because the amount of land is fixed, the supply is perfectly inelastic. Rapid increases in demand together with inelastic supply lead to large increases in the equilibrium rents on land so that economic growth makes rich landowners even richer.

Now consider the incidence of a tax on land. As we first saw in Chapter 6, the burden of a tax falls more heavily on the side of the market that is less elastic. A tax on land takes this principle to an extreme. Because the elasticity of supply is zero, the landowners bear the entire burden of the tax.

Consider next the question of efficiency. As we just discussed, the deadweight loss of a tax depends on the elasticities of supply and demand. Again, a tax on land is an extreme case. Because supply is perfectly inelastic, a tax on land does not alter the market allocation. There is no deadweight loss, and the government’s tax revenue exactly equals the loss of the landowners.

Although taxing land may look attractive in theory, it is not as straightforward in practice as it may appear. For a tax on land not to distort economic incentives, it must be a tax on raw land. Yet the value of land often comes from improvements, such as clearing trees, providing sewers, and building roads. Unlike the supply of raw land, the supply of improvements has an elasticity greater than zero. If a land tax were imposed on improvements, it would distort incentives. Landowners would respond by devoting fewer resources to improving their land.

Today, few economists support George’s proposal for a single tax on land. Not only is taxing improvements a potential problem, but the tax would not raise enough revenue to pay for the much larger government we have today. Yet many of George’s arguments remain valid. Here is the assessment of the eminent economist Milton Friedman a century after George’s book: “In my opinion, the least bad tax is the property tax on the unimproved value of land, the Henry George argument of many, many years ago.”

Contrast, panel (e) shows that tax revenue first rises with the size of the tax, but as the tax gets larger, the market shrinks so much that tax revenue starts to fall.

**Case Study**

One day in 1974, economist Arthur Laffer sat in a Washington restaurant with some prominent journalists and politicians. He took out a napkin and drew a figure on it to show how tax rates affect tax revenue. It looked much like panel (e) of our Figure 6. Laffer then suggested that the United States was on the
Tax rates affect work effort. This proposition helps explain why the U.S. economy differs from many others around the world.

U.S. Could Follow Europe’s High-Tax Path

Americans owe their economic edge over Europeans in part to the fact that they work more, a distinction often attributed to cultural differences: Americans want to consume more, while Europeans enjoy their leisure more.

As late as the 1970s, though, the French actually worked longer than Americans. The reason they now work one-third fewer hours has less to do with a yearning for the good life than it does with escalating taxes, including payroll taxes, in Europe. But Americans can’t afford to be smug: The U.S. may be headed in the same high-tax direction if it doesn’t tackle the looming crisis in Social Security and Medicare.

Edward Prescott of the University of Minnesota says Europe’s higher taxes made it more expensive to hire labor, even though take-home pay may not have increased much. The bigger the burden, the harder it is for employers to pay a salary that will entice someone to take a job rather than stay on public assistance, go to school, or retire early. Between the early 1970s and mid-1990s, he says, the French tax rate rose to 59 percent from 49 percent, while the U.S. tax rate held at 40 percent.

The result: The average French person of working age logged 24.4 hours a week in the early 1970s, one hour more than an American. By the mid-1990s, the French workweek had shrunk to 17.5 hours, while the U.S. workweek had grown to 25.9 hours.

Who Works Hardest?

In countries with higher taxes, people tend to work less.

<table>
<thead>
<tr>
<th>Country</th>
<th>Tax Rate</th>
<th>Workweek</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy</td>
<td>64%</td>
<td>16.5 hours</td>
</tr>
<tr>
<td>France</td>
<td>59%</td>
<td>17.5</td>
</tr>
<tr>
<td>Germany</td>
<td>59%</td>
<td>19.3</td>
</tr>
<tr>
<td>Canada</td>
<td>52%</td>
<td>22.8</td>
</tr>
<tr>
<td>U.K.</td>
<td>44%</td>
<td>22.9</td>
</tr>
<tr>
<td>U.S.</td>
<td>40%</td>
<td>25.9</td>
</tr>
<tr>
<td>Japan</td>
<td>37%</td>
<td>27.0</td>
</tr>
</tbody>
</table>

The relationship between work and tax rates was similar for the seven major industrial countries. The Japanese, with even lower taxes than the U.S., work more, and the Italians, with the highest taxes, work the least. The difference in hours was narrower in the 1970s, when the difference in tax rates was smaller.

Europe’s larger lesson for the U.S. may be about the costs of failing to prepare for the expense of the baby boomers’ retirement. The White House budget office says Social Security and Medicare have promised to pay out $18 trillion more than they will receive in revenue in coming decades. Closing that gap without any cuts in benefits would require a 7.1 percentage point increase in the combined Social Security–Medicare payroll tax, now at 15.3 percent.

“People would just stop working,” says Arthur Rolnick, research director of the Minneapolis Fed. As the work force shrunk, taxes would have to go up even more for the remaining workers.

Alan Auerbach of the University of California at Berkeley says the system’s generosity will have to be curtailed and “the sooner, the better.” Otherwise, American work habits again could look like those of the French.

evidence for Laffer’s view that U.S. tax rates had in fact reached such extreme levels.

Nonetheless, the Laffer curve (as it became known) captured the imagination of Ronald Reagan. David Stockman, budget director in the first Reagan administration, offers the following story:

[Reagan] had once been on the Laffer curve himself. “I came into the Big Money making pictures during World War II,” he would always say. At that time the wartime income surtax hit 90 percent. “You could only make four pictures and then you were in the top bracket,” he would continue. “So we all quit working after four pictures and went off to the country.” High tax rates caused less work. Low tax rates caused more. His experience proved it.

When Reagan ran for president in 1980, he made cutting taxes part of his platform. Reagan argued that taxes were so high that they were discouraging hard work. He argued that lower taxes would give people the proper incentive to work, which would raise economic well-being and perhaps even tax revenue. Because the cut in tax rates was intended to encourage people to increase the quantity of labor they supplied, the views of Laffer and Reagan became known as supply-side economics.

Economists continue to debate Laffer’s argument. Many believe that subsequent history refuted Laffer’s conjecture that lower tax rates would raise tax revenue. Yet because history is open to alternative interpretations, other economists view the events of the 1980s as more favorable to the supply-siders. To evaluate Laffer’s hypothesis definitively, we would need to rerun history without the Reagan tax cuts and see if tax revenues were higher or lower. Unfortunately, that experiment is impossible.

Some economists take an intermediate position on this issue. They believe that while an overall cut in tax rates normally reduces revenue, some taxpayers at some times may find themselves on the wrong side of the Laffer curve. Other things equal, a tax cut is more likely to raise tax revenue if the cut applies to those taxpayers facing the highest tax rates. In addition, Laffer’s argument may be more compelling when considering countries with much higher tax rates than the United States. In Sweden in the early 1980s, for instance, the typical worker faced a marginal tax rate of about 80 percent. Such a high tax rate provides a substantial disincentive to work. Studies have suggested that Sweden would indeed have raised more tax revenue if it had lowered its tax rates.

Economists disagree about these issues in part because there is no consensus about the size of the relevant elasticities. The more elastic that supply and demand are in any market, the more taxes in that market distort behavior, and the more likely it is that a tax cut will raise tax revenue. There is no debate, however, about the general lesson: How much revenue the government gains or loses from a tax change cannot be computed just by looking at tax rates. It also depends on how the tax change affects people’s behavior.

**QUICK QUIZ** If the government doubles the tax on gasoline, can you be sure that revenue from the gasoline tax will rise? Can you be sure that the deadweight loss from the gasoline tax will rise? Explain.
CONCLUSION

In this chapter we have used the tools developed in the previous chapter to further our understanding of taxes. One of the Ten Principles of Economics discussed in Chapter 1 is that markets are usually a good way to organize economic activity. In Chapter 7, we used the concepts of producer and consumer surplus to make this principle more precise. Here we have seen that when the government imposes taxes on buyers or sellers of a good, society loses some of the benefits of market efficiency. Taxes are costly to market participants not only because taxes transfer resources from those participants to the government but also because they alter incentives and distort market outcomes.

The analysis presented here and in Chapter 6 should give you a good basis for understanding the economic impact of taxes, but this is not the end of the story. Microeconomists study how best to design a tax system, including how to strike the right balance between equality and efficiency. Macroeconomists study how taxes influence the overall economy and how policymakers can use the tax system to stabilize economic activity and to achieve more rapid economic growth. So don’t be surprised that, as you continue your study of economics, the subject of taxation comes up yet again.

SUMMARY

- A tax on a good reduces the welfare of buyers and sellers of the good, and the reduction in consumer and producer surplus usually exceeds the revenue raised by the government. The fall in total surplus—the sum of consumer surplus, producer surplus, and tax revenue—is called the deadweight loss of the tax.

- Taxes have deadweight losses because they cause buyers to consume less and sellers to produce less, and these changes in behavior shrink the size of the market below the level that maximizes total surplus. Because the elasticities of supply and demand measure how much market participants respond to market conditions, larger elasticities imply larger deadweight losses.

- As a tax grows larger, it distorts incentives more, and its deadweight loss grows larger. Because a tax reduces the size of the market, however, tax revenue does not continually increase. It first rises with the size of a tax, but if a tax gets large enough, tax revenue starts to fall.
3. How do the elasticities of supply and demand affect the deadweight loss of a tax? Why do they have this effect?

4. Why do experts disagree about whether labor taxes have small or large deadweight losses?

5. What happens to the deadweight loss and tax revenue when a tax is increased?

1. What happens to consumer and producer surplus when the sale of a good is taxed? How does the change in consumer and producer surplus compare to the tax revenue? Explain.

2. Draw a supply-and-demand diagram with a tax on the sale of the good. Show the deadweight loss. Show the tax revenue.

**KEY CONCEPT**

deadweight loss, p. 163

**QUESTIONS FOR REVIEW**

1. What happens to consumer and producer surplus when the sale of a good is taxed? How does the change in consumer and producer surplus compare to the tax revenue? Explain.

2. Draw a supply-and-demand diagram with a tax on the sale of the good. Show the deadweight loss. Show the tax revenue.

**PROBLEMS AND APPLICATIONS**

1. The market for pizza is characterized by a downward-sloping demand curve and an upward-sloping supply curve.
   a. Draw the competitive market equilibrium. Label the price, quantity, consumer surplus, and producer surplus. Is there any deadweight loss? Explain.
   b. Suppose that the government forces each pizzeria to pay a $1 tax on each pizza sold. Illustrate the effect of this tax on the pizza market, being sure to label the consumer surplus, producer surplus, government revenue, and deadweight loss. How does each area compare to the pre-tax case?
   c. If the tax were removed, pizza eaters and sellers would be better off, but the government would lose tax revenue. Suppose that consumers and producers voluntarily transferred some of their gains to the government. Could all parties (including the government) be better off than they were with a tax? Explain using the labeled areas in your graph.

2. Evaluate the following two statements. Do you agree? Why or why not?
   a. “A tax that raises no revenue for the government cannot have any deadweight loss.”
   b. “A tax that raises no revenue for the government cannot have any deadweight loss.”

3. Consider the market for rubber bands.
   a. If this market has very elastic supply and very inelastic demand, how would the burden of a tax on rubber bands be shared between consumers and producers? Use the tools of consumer surplus and producer surplus in your answer.
   b. If this market has very inelastic supply and very elastic demand, how would the burden of a tax on rubber bands be shared between consumers and producers? Contrast your answer with your answer to part (a).

4. The 19th-century economist Henry George argued that the government should levy a sizable tax on land, the supply of which he took to be completely inelastic.
   a. George believed that economic growth increased the demand for land and made rich landowners richer at the expense of the tenants who made up the demand side of the market. Show this argument on an appropriately labeled diagram.
   b. Who bears the burden of a tax on land—the owners of land or the tenants on the land? Explain.
c. Is the deadweight loss of this tax large or small? Explain.
d. Many cities and towns today levy taxes on the value of real estate. Why might the above analysis of George’s land tax not apply to this modern tax?

5. Suppose that the government imposes a tax on heating oil.
   a. Would the deadweight loss from this tax likely be greater in the first year after it is imposed or in the fifth year? Explain.
   b. Would the revenue collected from this tax likely be greater in the first year after it is imposed or in the fifth year? Explain.

6. After economics class one day, your friend suggests that taxing food would be a good way to raise revenue because the demand for food is quite inelastic. In what sense is taxing food a “good” way to raise revenue? In what sense is it not a “good” way to raise revenue?

7. Daniel Patrick Moynihan, the late senator from New York, once introduced a bill that would levy a 10,000 percent tax on certain hollow-tipped bullets.
   a. Do you expect that this tax would raise much revenue? Why or why not?
   b. Even if the tax would raise no revenue, why might Senator Moynihan have proposed it?

8. The government places a tax on the purchase of socks.
   a. Illustrate the effect of this tax on equilibrium price and quantity in the sock market. Identify the following areas both before and after the imposition of the tax: total spending by consumers, total revenue for producers, and government tax revenue.
   b. Does the price received by producers rise or fall? Can you tell whether total receipts for producers rise or fall? Explain.
   c. Does the price paid by consumers rise or fall? Can you tell whether total spending by consumers rises or falls? Explain carefully. (Hint: Think about elasticity.) If total consumer spending falls, does consumer surplus rise? Explain.

9. Suppose the government currently raises $100 million through a 1-cent tax on widgets, and another $100 million through a 10-cent tax on gadgets. If the government doubled the tax rate on widgets and eliminated the tax on gadgets, would it raise more money than today, less money, or the same amount of money? Explain.

10. This chapter analyzed the welfare effects of a tax on a good. Consider now the opposite policy. Suppose that the government subsidizes a good: For each unit of the good sold, the government pays $2 to the buyer. How does the subsidy affect consumer surplus, producer surplus, tax revenue, and total surplus? Does a subsidy lead to a deadweight loss? Explain.

11. Hotel rooms in Smalltown go for $100, and 1,000 rooms are rented on a typical day.
   a. To raise revenue, the mayor decides to charge hotels a tax of $10 per rented room. After the tax is imposed, the going rate for hotel rooms rises to $108, and the number of rooms rented falls to 900. Calculate the amount of revenue this tax raises for Smalltown and the deadweight loss of the tax. (Hint: The area of a triangle is \( \frac{1}{2} \times \text{base} \times \text{height} \).)
   b. The mayor now doubles the tax to $20. The price rises to $116, and the number of rooms rented falls to 800. Calculate tax revenue and deadweight loss with this larger tax. Do they double, more than double, or less than double? Explain.

12. Suppose that a market is described by the following supply and demand equations:

\[
\begin{align*}
Q^s &= 2P \\
Q^d &= 300 - P
\end{align*}
\]

   a. Solve for the equilibrium price and the equilibrium quantity.
   b. Suppose that a tax of \( T \) is placed on buyers, so the new demand equation is

\[
Q^d = 300 - (P + T).
\]

Solve for the new equilibrium. What happens to the price received by sellers, the price paid by buyers, and the quantity sold?
c. Tax revenue is $T \times Q$. Use your answer to part (b) to solve for tax revenue as a function of $T$. Graph this relationship for $T$ between 0 and 300.

d. The deadweight loss of a tax is the area of the triangle between the supply and demand curves. Recalling that the area of a triangle is $\frac{1}{2} \times \text{base} \times \text{height}$, solve for deadweight loss as a function of $T$. Graph this relationship for $T$ between 0 and 300. (Hint: Looking sideways, the base of the deadweight loss triangle is $T$, and the height is the difference between the quantity sold with the tax and the quantity sold without the tax.)

e. The government now levies a tax on this good of $200 per unit. Is this a good policy? Why or why not? Can you propose a better policy?
This page intentionally left blank
Application: International Trade

If you check the labels on the clothes you are now wearing, you will probably find that some of your clothes were made in another country. A century ago, the textile and clothing industry was a major part of the U.S. economy, but that is no longer the case. Faced with foreign competitors that can produce quality goods at low cost, many U.S. firms have found it increasingly difficult to produce and sell textiles and clothing at a profit. As a result, they have laid off their workers and shut down their factories. Today, much of the textiles and clothing that Americans consume are imported.

The story of the textile industry raises important questions for economic policy: How does international trade affect economic well-being? Who gains and who loses from free trade among countries, and how do the gains compare to the losses?

Chapter 3 introduced the study of international trade by applying the principle of comparative advantage. According to this principle, all countries can benefit from trading with one another because trade allows each country to specialize in doing what it does best. But the analysis in Chapter 3 was incomplete. It did not
explain how the international marketplace achieves these gains from trade or how
the gains are distributed among various economic participants.

We now return to the study of international trade and take up these questions.
Over the past several chapters, we have developed many tools for analyzing how
markets work: supply, demand, equilibrium, consumer surplus, producer sur-
plus, and so on. With these tools, we can learn more about how international trade
affects economic well-being.

THE DETERMINANTS OF TRADE

Consider the market for textiles. The textile market is well suited to examining
the gains and losses from international trade: Textiles are made in many countries
around the world, and there is much world trade in textiles. Moreover, the textile
market is one in which policymakers often consider (and sometimes implement)
trade restrictions to protect domestic producers from foreign competitors. We
examine here the textile market in the imaginary country of Isoland.

THE EQUILIBRIUM WITHOUT TRADE

As our story begins, the Isolandian textile market is isolated from the rest of the
world. By government decree, no one in Isoland is allowed to import or export
textiles, and the penalty for violating the decree is so large that no one dares try.

Because there is no international trade, the market for textiles in Isoland con-
sists solely of Isolandian buyers and sellers. As Figure 1 shows, the domestic price
adjusts to balance the quantity supplied by domestic sellers and the quantity
demanded by domestic buyers. The figure shows the consumer and producer surplus
in the equilibrium without trade. The sum of consumer and producer surplus

---

**FIGURE 1**

The Equilibrium without International Trade

When an economy cannot trade in world markets, the price adjusts to balance domestic supply and demand. This figure shows consumer and producer surplus in an equilibrium without international trade for the textile market in the imaginary country of Isoland.
measures the total benefits that buyers and sellers receive from participating in the textile market.

Now suppose that, in an election upset, Isoland elects a new president. The president campaigned on a platform of “change” and promised the voters bold new ideas. Her first act is to assemble a team of economists to evaluate Isoladian trade policy. She asks them to report on three questions:

- If the government allows Isoladians to import and export textiles, what will happen to the price of textiles and the quantity of textiles sold in the domestic textile market?
- Who will gain from free trade in textiles and who will lose, and will the gains exceed the losses?
- Should a tariff (a tax on textile imports) be part of the new trade policy?

After reviewing supply and demand in their favorite textbook (this one, of course), the Isoladian economics team begins its analysis.

**THE WORLD PRICE AND COMPARATIVE ADVANTAGE**

The first issue our economists take up is whether Isoland is likely to become a textile importer or a textile exporter. In other words, if free trade is allowed, will Isoladians end up buying or selling textiles in world markets?

To answer this question, the economists compare the current Isoladian price of textiles to the price of textiles in other countries. We call the price prevailing in world markets the **world price**. If the world price of textiles is higher than the domestic price, then Isoland will export textiles once trade is permitted. Isoladian textile producers will be eager to receive the higher prices available abroad and will start selling their textiles to buyers in other countries. Conversely, if the world price of textiles is lower than the domestic price, then Isoland will import textiles. Because foreign sellers offer a better price, Isoladian textile consumers will quickly start buying textiles from other countries.

In essence, comparing the world price and the domestic price before trade indicates whether Isoland has a comparative advantage in producing textiles. The domestic price reflects the opportunity cost of textiles: It tells us how much an Isolodian must give up to obtain one unit of textiles. If the domestic price is low, the cost of producing textiles in Isoland is low, suggesting that Isoland has a comparative advantage in producing textiles relative to the rest of the world. If the domestic price is high, then the cost of producing textiles in Isoland is high, suggesting that foreign countries have a comparative advantage in producing textiles.

As we saw in Chapter 3, trade among nations is ultimately based on comparative advantage. That is, trade is beneficial because it allows each nation to specialize in doing what it does best. By comparing the world price and the domestic price before trade, we can determine whether Isoland is better or worse at producing textiles than the rest of the world.

**QUICK QUIZ** The country Autarka does not allow international trade. In Autarka, you can buy a wool suit for 3 ounces of gold. Meanwhile, in neighboring countries, you can buy the same suit for 2 ounces of gold. If Autarka were to allow free trade, would it import or export wool suits? Why?
To analyze the welfare effects of free trade, the Isolandian economists begin with the assumption that Isoland is a small economy compared to the rest of the world. This small-economy assumption means that Isoland’s actions have little effect on world markets. Specifically, any change in Isoland’s trade policy will not affect the world price of textiles. The Isolandians are said to be *price takers* in the world economy. That is, they take the world price of textiles as given. Isoland can be an exporting country by selling textiles at this price or an importing country by buying textiles at this price.

The small-economy assumption is not necessary to analyze the gains and losses from international trade. But the Isolandian economists know from experience (and from reading Chapter 2 of this book) that making simplifying assumptions is a key part of building a useful economic model. The assumption that Isoland is a small economy simplifies the analysis, and the basic lessons do not change in the more complicated case of a large economy.

**The Gains and Losses of an Exporting Country**

Figure 2 shows the Isolandian textile market when the domestic equilibrium price before trade is below the world price. Once trade is allowed, the domestic price rises to equal the world price. No seller of textiles would accept less than the world price, and no buyer would pay more than the world price.

---

**The Winners and Losers from Trade**

Once trade is allowed, the domestic price rises to equal the world price. The supply curve shows the quantity of textiles produced domestically, and the demand curve shows the quantity consumed domestically. Exports from Isoland equal the difference between the domestic quantity supplied and the domestic quantity demanded at the world price. Sellers are better off (producer surplus rises from C to B + C + D), and buyers are worse off (consumer surplus falls from A + B to A). Total surplus rises by an amount equal to area D, indicating that trade raises the economic well-being of the country as a whole.

---

**2 Figure**

**International Trade in an Exporting Country**

Once trade is allowed, the domestic price rises to equal the world price. The supply curve shows the quantity of textiles produced domestically, and the demand curve shows the quantity consumed domestically. Exports from Isoland equal the difference between the domestic quantity supplied and the domestic quantity demanded at the world price. Sellers are better off (producer surplus rises from C to B + C + D), and buyers are worse off (consumer surplus falls from A + B to A). Total surplus rises by an amount equal to area D, indicating that trade raises the economic well-being of the country as a whole.

---

**2 Figure**

**International Trade in an Exporting Country**

Once trade is allowed, the domestic price rises to equal the world price. The supply curve shows the quantity of textiles produced domestically, and the demand curve shows the quantity consumed domestically. Exports from Isoland equal the difference between the domestic quantity supplied and the domestic quantity demanded at the world price. Sellers are better off (producer surplus rises from C to B + C + D), and buyers are worse off (consumer surplus falls from A + B to A). Total surplus rises by an amount equal to area D, indicating that trade raises the economic well-being of the country as a whole.

---

**The Winners and Losers from Trade**

Once trade is allowed, the domestic price rises to equal the world price. The supply curve shows the quantity of textiles produced domestically, and the demand curve shows the quantity consumed domestically. Exports from Isoland equal the difference between the domestic quantity supplied and the domestic quantity demanded at the world price. Sellers are better off (producer surplus rises from C to B + C + D), and buyers are worse off (consumer surplus falls from A + B to A). Total surplus rises by an amount equal to area D, indicating that trade raises the economic well-being of the country as a whole.

---

**The Gains and Losses of an Exporting Country**

Figure 2 shows the Isolandian textile market when the domestic equilibrium price before trade is below the world price. Once trade is allowed, the domestic price rises to equal the world price. No seller of textiles would accept less than the world price, and no buyer would pay more than the world price.
After the domestic price has risen to equal the world price, the domestic quantity supplied differs from the domestic quantity demanded. The supply curve shows the quantity of textiles supplied by Isolandian sellers. The demand curve shows the quantity of textiles demanded by Isolandian buyers. Because the domestic quantity supplied is greater than the domestic quantity demanded, Isoland sells textiles to other countries. Thus, Isoland becomes a textile exporter.

Although domestic quantity supplied and domestic quantity demanded differ, the textile market is still in equilibrium because there is now another participant in the market: the rest of the world. One can view the horizontal line at the world price as representing the rest of the world’s demand for textiles. This demand curve is perfectly elastic because Isoland, as a small economy, can sell as many textiles as it wants at the world price.

Now consider the gains and losses from opening up trade. Clearly, not everyone benefits. Trade forces the domestic price to rise to the world price. Domestic producers of textiles are better off because they can now sell textiles at a higher price, but domestic consumers of textiles are worse off because they have to buy textiles at a higher price.

To measure these gains and losses, we look at the changes in consumer and producer surplus. Before trade is allowed, the price of textiles adjusts to balance domestic supply and domestic demand. Consumer surplus, the area between the demand curve and the before-trade price, is area \( A + B \). Producer surplus, the area between the supply curve and the before-trade price, is area \( C \). Total surplus before trade, the sum of consumer and producer surplus, is area \( A + B + C \).

After trade is allowed, the domestic price rises to the world price. Consumer surplus is reduced to area \( A \) (the area between the demand curve and the world price). Producer surplus is increased to area \( B + C + D \) (the area between the supply curve and the world price). Thus, total surplus with trade is area \( A + B + C + D \).

These welfare calculations show who wins and who loses from trade in an exporting country. Sellers benefit because producer surplus increases by the area \( B + D \). Buyers are worse off because consumer surplus decreases by the area \( B \). Because the gains of sellers exceed the losses of buyers by the area \( D \), total surplus in Isoland increases.

This analysis of an exporting country yields two conclusions:

- When a country allows trade and becomes an exporter of a good, domestic producers of the good are better off, and domestic consumers of the good are worse off.
- Trade raises the economic well-being of a nation in the sense that the gains of the winners exceed the losses of the losers.

**The Gains and Losses of an Importing Country**

Now suppose that the domestic price before trade is above the world price. Once again, after trade is allowed, the domestic price must equal the world price. As Figure 3 shows, the domestic quantity supplied is less than the domestic quantity demanded. The difference between the domestic quantity demanded and the domestic quantity supplied is bought from other countries, and Isoland becomes a textile importer.

In this case, the horizontal line at the world price represents the supply of the rest of the world. This supply curve is perfectly elastic because Isoland is a small economy and, therefore, can buy as many textiles as it wants at the world price.
Now consider the gains and losses from trade. Once again, not everyone benefits. When trade forces the domestic price to fall, domestic consumers are better off (they can now buy textiles at a lower price), and domestic producers are worse off (they now have to sell textiles at a lower price). Changes in consumer and producer surplus measure the size of the gains and losses. Before trade, consumer surplus is area $A$, producer surplus is area $B + C$, and total surplus is area $A + B + C$. After trade is allowed, consumer surplus is area $A + B + D$, producer surplus is area $C$, and total surplus is area $A + B + C + D$.

These welfare calculations show who wins and who loses from trade in an importing country. Buyers benefit because consumer surplus increases by the area $B + D$, and sellers are worse off because producer surplus falls by the area $B$. The gains of buyers exceed the losses of sellers, and total surplus increases by the area $D$.

This analysis of an importing country yields two conclusions parallel to those for an exporting country:

- When a country allows trade and becomes an importer of a good, domestic consumers of the good are better off, and domestic producers of the good are worse off.
- Trade raises the economic well-being of a nation in the sense that the gains of the winners exceed the losses of the losers.
Having completed our analysis of trade, we can better understand one of the 
Ten Principles of Economics in Chapter 1: Trade can make everyone better off. If Iso-
land opens its textile market to international trade, the change will create winners 
and losers, regardless of whether Isoland ends up exporting or importing textiles. 
In either case, however, the gains of the winners exceed the losses of the losers, so 
the winners could compensate the losers and still be better off. In this sense, trade can make everyone better off. But will trade make everyone better off? Probably 
not. In practice, compensation for the losers from international trade is rare. With-
out such compensation, opening an economy to international trade is a policy that 
expands the size of the economic pie, while perhaps leaving some participants in 
the economy with a smaller slice. 

We can now see why the debate over trade policy is often contentious. Whenever a policy creates winners and losers, the stage is set for a political battle. Nations sometimes fail to enjoy the gains from trade because the losers from free 
trade are better organized than the winners. The losers may turn their cohesive-
ness into political clout, lobbying for trade restrictions such as tariffs or import 
quotas.

The Effects of a Tariff

The Isolandian economists next consider the effects of a tariff—a tax on imported 
goods. The economists quickly realize that a tariff on textiles will have no effect 
if Isoland becomes a textile exporter. If no one in Isoland is interested in import-
ing textiles, a tax on textile imports is irrelevant. The tariff matters only if Isoland 
becomes a textile importer. Concentrating their attention on this case, the econo-
mists compare welfare with and without the tariff.

Figure 4 shows the Isolandian market for textiles. Under free trade, the domestic 
price equals the world price. A tariff raises the price of imported textiles above the 
world price by the amount of the tariff. Domestic suppliers of textiles, who com-
pe with suppliers of imported textiles, can now sell their textiles for the world 
price plus the amount of the tariff. Thus, the price of textiles—both imported and 
domestic—rises by the amount of the tariff and is, therefore, closer to the price 
that would prevail without trade.

The change in price affects the behavior of domestic buyers and sellers. Because 
the tariff raises the price of textiles, it reduces the domestic quantity demanded 
from \( Q_1^D \) to \( Q_2^D \) and raises the domestic quantity supplied from \( Q_1^S \) to \( Q_2^S \). Thus, the tariff reduces the quantity of imports and moves the domestic market closer to its equilib-
rium without trade.

Now consider the gains and losses from the tariff. Because the tariff raises the 
domestic price, domestic sellers are better off, and domestic buyers are worse off. 
In addition, the government raises revenue. To measure these gains and losses, 
we look at the changes in consumer surplus, producer surplus, and government 
revenue. These changes are summarized in the table in Figure 4.

Before the tariff, the domestic price equals the world price. Consumer surplus, 
the area between the demand curve and the world price, is area \( A + B + C + D + E + F \). Producer surplus, the area between the supply curve and the world price, 
is area \( G \). Government revenue equals zero. Total surplus, the sum of consumer 
surplus, producer surplus, and government revenue, is area \( A + B + C + D + E + F + G \).

Once the government imposes a tariff, the domestic price exceeds the world 
price by the amount of the tariff. Consumer surplus is now area \( A + B \). Producer
surplus is area $C + G$. Government revenue, which is the quantity of after-tariff imports times the size of the tariff, is the area $E$. Thus, total surplus with the tariff is area $A + B + C + D + E + G$.

To determine the total welfare effects of the tariff, we add the change in consumer surplus (which is negative), the change in producer surplus (positive), and the change in government revenue (positive). We find that total surplus in the market decreases by the area $D + F$. This fall in total surplus is called the deadweight loss of the tariff.

A tariff causes a deadweight loss simply because a tariff is a type of tax. Like most taxes, it distorts incentives and pushes the allocation of scarce resources
away from the optimum. In this case, we can identify two effects. First, when the tariff raises the domestic price of textiles above the world price, it encourages domestic producers to increase production from $Q_1^S$ to $Q_2^S$. Even though the cost of making these incremental units exceeds the cost of buying them at the world price, the tariff makes it profitable for domestic producers to manufacture them nonetheless. Second, when the tariff raises the price that domestic textile consumers have to pay, it encourages them to reduce consumption of textiles from $Q_1^D$ to $Q_2^D$. Even though domestic consumers value these incremental units at more than the world price, the tariff induces them to cut back their purchases. Area D represents the deadweight loss from the overproduction of textiles, and area F represents the deadweight loss from the underconsumption. The total deadweight loss of the tariff is the sum of these two triangles.

**The Lessons for Trade Policy**

The team of Isolandian economists can now write to the new president:

Dear Madame President,

You asked us three questions about opening up trade. After much hard work, we have the answers.


**Question:** If the government allows Isolandians to import and export textiles, what will happen to the price of textiles and the quantity of textiles sold in the domestic textile market?

**Answer:** Once trade is allowed, the Isolandian price of textiles will be driven to equal the price prevailing around the world.

If the world price is now higher than the Isolandian price, our price will rise. The higher price will reduce the amount of textiles Isolandians consume and raise the amount of textiles that Isolandians produce. Isoland will, therefore, become a textile exporter. This occurs because, in this case, Isoland has a comparative advantage in producing textiles.

Conversely, if the world price is now lower than the Isolandian price, our price will fall. The lower price will raise the amount of textiles that Isolandians consume and lower the amount of textiles that Isolandians produce. Isoland will, therefore, become a textile importer. This occurs because, in this case, other countries have a comparative advantage in producing textiles.

**Question:** Who will gain from free trade in textiles and who will lose, and will the gains exceed the losses?

**Answer:** The answer depends on whether the price rises or falls when trade is allowed. If the price rises, producers of textiles gain, and consumers of textiles lose. If the price falls, consumers gain, and producers lose. In both cases, the gains are larger than the losses. Thus, free trade raises the total welfare of Isolandians.

**Question:** Should a tariff be part of the new trade policy?

**Answer:** A tariff has an impact only if Isoland becomes a textile importer. In this case, a tariff moves the economy closer to the no-trade equilibrium and, like most taxes, has deadweight losses. Although a tariff improves the welfare of domestic producers and raises revenue for the government, these gains are more than offset by the losses suffered by consumers. The best policy, from the standpoint of economic efficiency, would be to allow trade without a tariff.

We hope you find these answers helpful as you decide on your new policy.

Your faithful servants,

Isolandian economics team

**Other Benefits of International Trade**

The conclusions of the Isolandian economics team are based on the standard analysis of international trade. Their analysis uses the most fundamental tools in the economist’s toolbox: supply, demand, and producer and consumer surplus. It shows that there are winners and losers when a nation opens itself up to trade, but the gains to the winners exceed the losses of the losers.

The case for free trade can be made even stronger, however, because there are several other economic benefits of trade beyond those emphasized in the standard analysis. Here, in a nutshell, are some of these other benefits:

- **Increased variety of goods.** Goods produced in different countries are not exactly the same. German beer, for instance, is not the same as American beer. Free trade gives consumers in all countries greater variety from which to choose.
Should the Winners from Free Trade Compensate the Losers?

Political candidates often say that the government should help those made worse off by international trade. In this opinion piece, an economist makes the opposite case.

What to Expect When You’re Free Trading

By Steven E. Landsburg

In the days before Tuesday’s Republican presidential primary in Michigan, Mitt Romney and John McCain battled over what the government owes to workers who lose their jobs because of the foreign competition unleashed by free trade. Their rhetoric differed—Mr. Romney said he would “fight for every single job,” while Mr. McCain said some jobs “are not coming back”—but their proposed policies were remarkably similar: educate and retrain the workers for new jobs.

All economists know that when American jobs are outsourced, Americans as a group are net winners. What we lose through lower wages is more than offset by what we gain through lower prices. In other words, the winners can more than afford to compensate the losers. Does that mean they ought to? Does it create a moral mandate for the taxpayer-subsidized retraining programs proposed by Mr. McCain and Mr. Romney?

Um, no. Even if you’ve just lost your job, there’s something fundamentally churlish about blaming the very phenomenon that’s elevated you above the subsistence level since the day you were born. If the world owes you compensation for enduring the downside of trade, what do you owe the world for enjoying the upside?

I doubt there’s a human being on earth who hasn’t benefited from the opportunity to trade freely with his neighbors. Imagine what your life would be like if you had to grow your own food, make your own clothes and rely on your grandmother’s home remedies for health care. Access to a trained physician might reduce the demand for grandma’s home remedies, but—especially at her age—she’s still got plenty of reason to be thankful for having a doctor.

Some people suggest, however, that it makes sense to isolate the moral effects of a single new trading opportunity or free trade agreement. Surely we have fellow citizens who are hurt by those agreements, at least in the limited sense that they’d be better off in a world where trade flourishes, except in this one instance. What do we owe those fellow citizens?

One way to think about that is to ask what your moral instincts tell you in analogous situations. Suppose, after years of buying shampoo at your local pharmacy, you discover you can order the same shampoo for less money on the Web. Do you have an obligation to compensate your pharmacist? If you move to a cheaper apartment, should you compensate your landlord? When you eat at McDonald’s, should you compensate the owners of the diner next door? Public policy should not be designed to advance moral instincts that we all reject every day of our lives.

In what morally relevant way, then, might displaced workers differ from displaced pharmacists or displaced landlords? You might argue that pharmacists and landlords have always faced cutthroat competition and therefore knew what they were getting into, while decades of tariffs and quotas have led manufacturing workers to expect a modicum of protection. That expectation led them to develop certain skills, and now it’s unfair to pull the rug out from under them.

Once again, that argument does not mesh with our everyday instincts. For many decades, schoolyard bullying has been a profitable occupation. All across America, bullies have built up skills so they can take advantage of that opportunity. If we toughen the rules to make bullying unprofitable, must we compensate the bullies?

Bullying and protectionism have a lot in common. They both use force (either directly or through the power of the law) to enrich someone else at your involuntary expense. If you’re forced to pay $20 an hour to an American for goods you could have bought from a Mexican for $5 an hour, you’re being extorted. When a free trade agreement allows you to buy from the Mexican after all, rejoice in your liberation—even if Mr. McCain, Mr. Romney and the rest of the presidential candidates don’t want you to.

• **Lower costs through economies of scale.** Some goods can be produced at low cost only if they are produced in large quantities—a phenomenon called *economies of scale*. A firm in a small country cannot take full advantage of economies of scale if it can sell only in a small domestic market. Free trade gives firms access to larger world markets and allows them to realize economies of scale more fully.

• **Increased competition.** A company shielded from foreign competitors is more likely to have market power, which in turn gives it the ability to raise prices above competitive levels. This is a type of market failure. Opening up trade fosters competition and gives the invisible hand a better chance to work its magic.

• **Enhanced flow of ideas.** The transfer of technological advances around the world is often thought to be linked to the trading of the goods that embody those advances. The best way for a poor agricultural nation to learn about the computer revolution, for instance, is to buy some computers from abroad rather than trying to make them domestically.

Thus, free international trade increases variety for consumers, allows firms to take advantage of economies of scale, makes markets more competitive, and facilitates the spread of technology. If the Isolandian economists also took these effects into account, their advice to their president would be even more forceful.

**Quick Quiz**

Draw a supply and demand diagram for wool suits in the country of Autarka. When trade is allowed, the price of a suit falls from 3 to 2 ounces of gold. In your diagram, show the change in consumer surplus, the change in producer surplus, and the change in total surplus. How would a tariff on suit imports alter these effects?

---

**THE ARGUMENTS FOR RESTRICTING TRADE**

The letter from the economics team starts to persuade the new president of Isoland to consider allowing trade in textiles. She notes that the domestic price is now high compared to the world price. Free trade would, therefore, cause the price of textiles to fall and hurt domestic textiles producers. Before implementing the new policy, she asks Isolandian textile companies to comment on the economists’ advice.

Not surprisingly, the textile companies oppose free trade in textiles. They believe that the government should protect the domestic textile industry from foreign competition. Let’s consider some of the arguments they might give to support their position and how the economics team would respond.

**The Jobs Argument**

Opponents of free trade often argue that trade with other countries destroys domestic jobs. In our example, free trade in textiles would cause the price of textiles to fall, reducing the quantity of textiles produced in Isoland and thus reducing employment in the Isolandian textile industry. Some Isolandian textile workers would lose their jobs.
Offshore Outsourcing

If you buy a new computer and call the company for tech support, you shouldn’t be surprised if you end up talking to someone in Bangalore, India. In 2004, the author of this textbook, while an adviser to President Bush, was asked about the movement of such jobs overseas. I replied that the trend was “probably a plus for the economy in the long run.” Most economists agreed, but some elected officials responded differently.

The Economics of Progress
By George F. Will

It is difficult to say something perfectly, precisely false. But House Speaker Dennis Hastert did when participating in the bipartisan piling-on against the president’s economic adviser, who imprudently said something sensible.

John Kerry and John Edwards, who are not speaking under oath and who know that economic illiteracy has never been a disqualification for high office, have led the scrum against the chairman of the president’s Council of Economic Advisers, N. Gregory Mankiw, who said the arguments for free trade apply to trade in services as well as manufactured goods. But the prize for the pithiest nonsense went to Hastert: “An economy suffers when jobs disappear.”

So the economy suffered when automobiles caused the disappearance of the jobs of most blacksmiths, buggy makers, operators of livery stables, etc.? The economy did not seem to be suffering in 1999, when 33 million jobs were wiped out—by an economic dynamism that created 35.7 million jobs. How many of the 4,500 U.S. jobs that IBM is planning to create this year will be made possible by sending 3,000 jobs overseas?

Hastert’s ideal economy, where jobs do not disappear, existed almost everywhere for almost everyone through almost all of human history. In, say, 12th-century France, the ox behind which a man plowed a field changed, but otherwise the plowman was doing what generations of his ancestors had done and what generations of his descendants were to do. Those were the good old days, before economic growth . . . .

For the highly competent workforce of this wealthy nation, the loss of jobs is not a zero-sum game: It is a trading up in social rewards. When the presidential candidates were recently in South Carolina, histrionically lamenting the loss of textile jobs, they surely noticed the huge BMW presence. It is the “offshoring” of German jobs because Germany’s irrational labor laws, among other things, give America a comparative advantage. Such economic calculation explains the manufacture of Mercedes-Benzes in Alabama, Hondas in Ohio, Toyotas in California.

As long as the American jobs going offshore were blue-collar jobs, the political issue did not attain the heat it has now that white-collar job losses frighten a more articulate, assertive social class . . . .

Kerry says offshoring is done by “Benedict Arnold CEOs.” But if he wants to improve the health of U.S. airlines, and the security of the jobs and pensions of most airline employees, should he not applaud Delta for saving $25 million a year by sending some reservation services to India?

Does Kerry really want to restrain the rise of health care costs? Does he oppose having X-rays analyzed in India at a fraction of the U.S. cost?

In November, Indiana Gov. Joseph Kernan canceled a $15 million contract with a firm in India to process state unemployment claims. The contract was given to a U.S. firm that will charge $23 million. Because of this 53 percent price increase, there will be 8 million fewer state dollars for schools, hospitals, law enforcement, etc. And the benefit to Indiana is . . . what?

When Kernan made this gesture he probably was wearing something that was wholly or partly imported and that at one time, before offshoring, would have been entirely made here. Such potential embarrassments are among the perils of making moral grandstanding into an economic policy.

Yet free trade creates jobs at the same time that it destroys them. When Isolandians buy textiles from other countries, those countries obtain the resources to buy other goods from Isoland. Isolandian workers would move from the textile industry to those industries in which Isoland has a comparative advantage. The transition may impose hardship on some workers in the short run, but it allows Isolandians as a whole to enjoy a higher standard of living.

Opponents of trade are often skeptical that trade creates jobs. They might respond that everything can be produced more cheaply abroad. Under free trade, they might argue, Isolandians could not be profitably employed in any industry. As Chapter 3 explains, however, the gains from trade are based on comparative advantage, not absolute advantage. Even if one country is better than another country at producing everything, each country can still gain from trading with the other. Workers in each country will eventually find jobs in an industry in which that country has a comparative advantage.

**The National-Security Argument**

When an industry is threatened with competition from other countries, opponents of free trade often argue that the industry is vital for national security. For example, if Isoland were considering free trade in steel, domestic steel companies might point out that steel is used to make guns and tanks. Free trade would allow Isoland to become dependent on foreign countries to supply steel. If a war later broke out and the foreign supply was interrupted, Isoland might be unable to produce enough steel and weapons to defend itself.

Economists acknowledge that protecting key industries may be appropriate when there are legitimate concerns over national security. Yet they fear that this argument may be used too quickly by producers eager to gain at consumers’ expense.

One should be wary of the national-security argument when it is made by representatives of industry rather than the defense establishment. Companies have an incentive to exaggerate their role in national defense to obtain protection from foreign competition. A nation’s generals may see things very differently. Indeed, when the military is a consumer of an industry’s output, it would benefit from imports. Cheaper steel in Isoland, for example, would allow the Isolandian military to accumulate a stockpile of weapons at lower cost.

**The Infant-Industry Argument**

New industries sometimes argue for temporary trade restrictions to help them get started. After a period of protection, the argument goes, these industries will mature and be able to compete with foreign firms.

Similarly, older industries sometimes argue that they need temporary protection to help them adjust to new conditions. For example, in 2002, President Bush imposed temporary tariffs on imported steel. He said, “I decided that imports were severely affecting our industry, an important industry.” The tariff, which lasted 20 months, offered “temporary relief so that the industry could restructure itself.”

Economists are often skeptical about such claims, largely because the infant-industry argument is difficult to implement in practice. To apply protection suc-
cessfully, the government would need to decide which industries will eventually be profitable and decide whether the benefits of establishing these industries exceed the costs of this protection to consumers. Yet “picking winners” is extraordinarily difficult. It is made even more difficult by the political process, which often awards protection to those industries that are politically powerful. And once a powerful industry is protected from foreign competition, the “temporary” policy is sometimes hard to remove.

In addition, many economists are skeptical about the infant-industry argument in principle. Suppose, for instance, that an industry is young and unable to compete profitably against foreign rivals, but there is reason to believe that the industry can be profitable in the long run. In this case, firm owners should be willing to incur temporary losses to obtain the eventual profits. Protection is not necessary for an infant industry to grow. History shows that start-up firms often incur temporary losses and succeed in the long run, even without protection from competition.

**The Unfair-Competition Argument**

A common argument is that free trade is desirable only if all countries play by the same rules. If firms in different countries are subject to different laws and regulations, then it is unfair (the argument goes) to expect the firms to compete in the international marketplace. For instance, suppose that the government of Neighborland subsidizes its textile industry by giving textile companies large tax breaks. The Isolodian textile industry might argue that it should be protected from this foreign competition because Neighborland is not competing fairly.

Would it, in fact, hurt Isoland to buy textiles from another country at a subsidized price? Certainly, Isolodian textile producers would suffer, but Isolodian textile consumers would benefit from the low price. The case for free trade is no different: The gains of the consumers from buying at the low price would exceed the losses of the producers. Neighborland’s subsidy to its textile industry may be a bad policy, but it is the taxpayers of Neighborland who bear the burden. Isoland can benefit from the opportunity to buy textiles at a subsidized price.

**The Protection-as-a-Bargaining-Chip Argument**

Another argument for trade restrictions concerns the strategy of bargaining. Many policymakers claim to support free trade but, at the same time, argue that trade restrictions can be useful when we bargain with our trading partners. They claim that the threat of a trade restriction can help remove a trade restriction already imposed by a foreign government. For example, Isoland might threaten to impose a tariff on textiles unless Neighborland removes its tariff on wheat. If Neighborland responds to this threat by removing its tariff, the result can be freer trade.

The problem with this bargaining strategy is that the threat may not work. If it doesn’t work, the country faces a choice between two bad options. It can carry out its threat and implement the trade restriction, which would reduce its own economic welfare. Or it can back down from its threat, which would cause it to lose prestige in international affairs. Faced with this choice, the country would probably wish that it had never made the threat in the first place.
A country can take one of two approaches to achieving free trade. It can take a unilaterial approach and remove its trade restrictions on its own. This is the approach that Great Britain took in the 19th century and that Chile and South Korea have taken in recent years. Alternatively, a country can take a multilateral approach and reduce its trade restrictions while other countries do the same. In other words, it can bargain with its trading partners in an attempt to reduce trade restrictions around the world.

One important example of the multilateral approach is the North American Free Trade Agreement (NAFTA), which in 1993 lowered trade barriers among the United States, Mexico, and Canada. Another is the General Agreement on Tariffs and Trade (GATT), which is a continuing series of negotiations among many of the world’s countries with the goal of promoting free trade. The United States
helped to found GATT after World War II in response to the high tariffs imposed during the Great Depression of the 1930s. Many economists believe that the high tariffs contributed to the worldwide economic hardship of that period. GATT has successfully reduced the average tariff among member countries from about 40 percent after World War II to about 5 percent today.

The rules established under GATT are now enforced by an international institution called the World Trade Organization (WTO). The WTO was established in 1995 and has its headquarters in Geneva, Switzerland. As of July 2007, 151 countries have joined the organization, accounting for more than 97 percent of world trade. The functions of the WTO are to administer trade agreements, provide a forum for negotiations, and handle disputes among member countries.

What are the pros and cons of the multilateral approach to free trade? One advantage is that the multilateral approach has the potential to result in freer trade than a unilateral approach because it can reduce trade restrictions abroad as well as at home. If international negotiations fail, however, the result could be more restricted trade than under a unilateral approach.
In addition, the multilateral approach may have a political advantage. In most markets, producers are fewer and better organized than consumers—and thus wield greater political influence. Reducing the Isolandian tariff on textiles, for example, may be politically difficult if considered by itself. The textile companies would oppose free trade, and the buyers of textiles who would benefit are so numerous that organizing their support would be difficult. Yet suppose that Neighborland promises to reduce its tariff on wheat at the same time that Isoland reduces its tariff on textiles. In this case, the Isolandian wheat farmers, who are also politically powerful, would back the agreement. Thus, the multilateral approach to free trade can sometimes win political support when a unilateral approach cannot.

**Quick Quiz** The textile industry of Autarka advocates a ban on the import of wool suits. Describe five arguments its lobbyists might make. Give a response to each of these arguments.

---

**CONCLUSION**

Economists and the public often disagree about free trade. In December 2007, the *Los Angeles Times* asked the American public, “Generally speaking, do you believe that free international trade has helped or hurt the economy, or hasn’t it made a difference to the economy one way or the other?” Only 27 percent of those polled said free international trade helped, whereas 44 percent thought it hurt. (The rest thought it made no difference or were unsure.) By contrast, most economists support free international trade. They view free trade as a way of allocating production efficiently and raising living standards both at home and abroad.

Economists view the United States as an ongoing experiment that confirms the virtues of free trade. Throughout its history, the United States has allowed unrestricted trade among the states, and the country as a whole has benefited from the specialization that trade allows. Florida grows oranges, Texas pumps oil, California makes wine, and so on. Americans would not enjoy the high standard of living they do today if people could consume only those goods and services produced in their own states. The world could similarly benefit from free trade among countries.

To better understand economists’ view of trade, let’s continue our parable. Suppose that the president of Isoland, after reading the latest poll results, ignores the advice of her economics team and decides not to allow free trade in textiles. The country remains in the equilibrium without international trade.

Then, one day, some Isolandian inventor discovers a new way to make textiles at very low cost. The process is quite mysterious, however, and the inventor insists on keeping it a secret. What is odd is that the inventor doesn’t need traditional inputs such as cotton or wool. The only material input he needs is wheat. And even more oddly, to manufacture textiles from wheat, he hardly needs any labor input at all.

The inventor is hailed as a genius. Because everyone buys clothing, the lower cost of textiles allows all Isolanders to enjoy a higher standard of living. Workers who had previously produced textiles experience some hardship when their factories close, but eventually, they find work in other industries. Some become
farmers and grow the wheat that the inventor turns into textiles. Others enter new industries that emerge as a result of higher Isolandian living standards. Everyone understands that the displacement of workers in outmoded industries is an inevitable part of technological progress and economic growth.

After several years, a newspaper reporter decides to investigate this mysterious new textiles process. She sneaks into the inventor’s factory and learns that the inventor is a fraud. The inventor has not been making textiles at all. Instead, he has been smuggling wheat abroad in exchange for textiles from other countries. The only thing that the inventor had discovered was the gains from international trade.

When the truth is revealed, the government shuts down the inventor’s operation. The price of textiles rises, and workers return to jobs in textile factories. Living standards in Isoland fall back to their former levels. The inventor is jailed and held up to public ridicule. After all, he was no inventor. He was just an economist.

**SUMMARY**

- The effects of free trade can be determined by comparing the domestic price without trade to the world price. A low domestic price indicates that the country has a comparative advantage in producing the good and that the country will become an exporter. A high domestic price indicates that the rest of the world has a comparative advantage in producing the good and that the country will become an importer.

- When a country allows trade and becomes an exporter of a good, producers of the good are better off, and consumers of the good are worse off. When a country allows trade and becomes an importer of a good, consumers are better off, and producers are worse off. In both cases, the gains from trade exceed the losses.

- A tariff—a tax on imports—moves a market closer to the equilibrium that would exist without trade and, therefore, reduces the gains from trade. Although domestic producers are better off and the government raises revenue, the losses to consumers exceed these gains.

- There are various arguments for restricting trade: protecting jobs, defending national security, helping infant industries, preventing unfair competition, and responding to foreign trade restrictions. Although some of these arguments have some merit in some cases, economists believe that free trade is usually the better policy.
KEY CONCEPTS

world price, p. 179
tariff, p. 183

QUESTIONS FOR REVIEW

1. What does the domestic price that prevails without international trade tell us about a nation’s comparative advantage?
2. When does a country become an exporter of a good? An importer?
3. Draw the supply-and-demand diagram for an importing country. What is consumer surplus and producer surplus before trade is allowed? What is consumer surplus and producer surplus with free trade? What is the change in total surplus?
4. Describe what a tariff is and its economic effects.
5. List five arguments often given to support trade restrictions. How do economists respond to these arguments?
6. What is the difference between the unilateral and multilateral approaches to achieving free trade? Give an example of each.

PROBLEMS AND APPLICATIONS

1. Mexico represents a small part of the world orange market.
   a. Draw a diagram depicting the equilibrium in the Mexican orange market without international trade. Identify the equilibrium price, equilibrium quantity, consumer surplus, and producer surplus.
   b. Suppose that the world orange price is below the Mexican price before trade and that the Mexican orange market is now opened to trade. Identify the new equilibrium price, quantity consumed, quantity produced domestically, and quantity imported. Also show the change in the surplus of domestic consumers and producers. Has total surplus increased or decreased?

2. The world price of wine is below the price that would prevail in Canada in the absence of trade.
   a. Assuming that Canadian imports of wine are a small part of total world wine production, draw a graph for the Canadian market for wine under free trade. Identify consumer surplus, producer surplus, and total surplus in an appropriate table.
   b. Now suppose that an unusual shift of the Gulf Stream leads to an unseasonably cold summer in Europe, destroying much of the grape harvest there. What effect does this shock have on the world price of wine? Using your graph and table from part (a), show the effect on consumer surplus, producer sur-
6. Consider the arguments for restricting trade.
   a. Assume you are a lobbyist for timber, an established industry suffering from low-priced foreign competition. Which two or three of the five arguments do you think would be most persuasive to the average member of Congress as to why he or she should support trade restrictions? Explain your reasoning.
   b. Now assume you are an astute student of economics (hopefully not a hard assumption). Although all the arguments for restricting trade have their shortcomings, name the two or three arguments that seem to make the most economic sense to you. For each, describe the economic rationale for and against these arguments for trade restrictions.

7. Senator Ernest Hollings once wrote that “consumers do not benefit from lower-priced imports. Glance through some mail-order catalogs and you’ll see that consumers pay exactly the same price for clothing whether it is U.S.-made or imported.” Comment.

8. The nation of Textilia does not allow imports of clothing. In its equilibrium without trade, a T-shirt costs $20, and the equilibrium quantity is 3 million T-shirts. One day, after reading Adam Smith’s *The Wealth of Nations* while on vacation, the president decides to open the Textilian market to international trade. The market price of a T-shirt falls to the world price of $16. The number of T-shirts consumed in Textilia rises to 4 million, while the number of T-shirts produced declines to 1 million.
   a. Illustrate the situation just described in a graph. Your graph should show all the numbers.
   b. Calculate the change in consumer surplus, producer surplus, and total surplus that results from opening up trade. (Hint: Recall that the area of a triangle is \( \frac{1}{2} \times \text{base} \times \text{height} \).)
9. China is a major producer of grains, such as wheat, corn, and rice. In 2008 the Chinese government, concerned that grain exports were driving up food prices for domestic consumers, imposed a tax on grain exports.
   a. Draw the graph that describes the market for grain in an exporting country. Use this graph as the starting point to answer the following questions.
   b. How does an export tax affect domestic grain prices?
   c. How does it affect the welfare of domestic consumers, the welfare of domestic producers, and government revenue?
   d. What happens to total welfare in China, as measured by the sum of consumer surplus, producer surplus, and tax revenue?

10. Consider a country that imports a good from abroad. For each of following statements, say whether it is true or false. Explain your answer.
   a. “The greater the elasticity of demand, the greater the gains from trade.”
   b. “If demand is perfectly inelastic, there are no gains from trade.”
   c. “If demand is perfectly inelastic, consumers do not benefit from trade.”

11. Kawmin is a small country that produces and consumes jelly beans. The world price of jelly beans is $1 per bag, and Kawmin’s domestic demand and supply for jelly beans are governed by the following equations:
   \[ Q^D = 8 - P \]
   \[ Q^S = P, \]
   where \( P \) is in dollars per bag and \( Q \) is in bags of jelly beans.
   a. Draw a well-labeled graph of the situation in Kawmin if the nation does not allow trade. Calculate the following (recalling that the area of a triangle is \( \frac{1}{2} \times \text{base} \times \text{height} \)):
      - the equilibrium price and quantity,
      - consumer surplus,
      - producer surplus,
      - and total surplus.
   b. Kawmin then opens the market to trade. Draw another graph to describe the new situation in the jelly bean market. Calculate the equilibrium price, quantities of consumption and production, imports, consumer surplus, producer surplus, and total surplus.
   c. After awhile, the Czar of Kawmin responds to the pleas of jelly bean producers by placing a $1 per bag tariff on jelly bean imports. On a graph, show the effects of this tariff. Calculate the equilibrium price, quantities of consumption and production, imports, consumer surplus, producer surplus, government revenue, and total surplus.
   d. What are the gains from opening up trade? What are the deadweight losses from restricting trade with the tariff? Give numerical answers.

12. Assume the United States is an importer of televisions and there are no trade restrictions. U.S. consumers buy 1 million televisions per year, of which 400,000 are produced domestically and 600,000 are imported.
   a. Suppose that a technological advance among Japanese television manufacturers causes the world price of televisions to fall by $100. Draw a graph to show how this change affects the welfare of U.S. consumers and U.S. producers and how it affects total surplus in the United States.
   b. After the fall in price, consumers buy 1.2 million televisions, of which 200,000 are produced domestically and 1 million are imported. Calculate the change in consumer surplus, producer surplus, and total surplus from the price reduction.
   c. If the government responded by putting a $100 tariff on imported televisions, what would this do? Calculate the revenue that would be raised and the deadweight loss. Would it be a good policy from the standpoint of U.S. welfare? Who might support the policy?
d. Suppose that the fall in price is attributable not to technological advance but to a $100 per television subsidy from the Japanese government to Japanese industry. How would this affect your analysis?

13. Consider a small country that exports steel. Suppose that a “pro-trade” government decides to subsidize the export of steel by paying a certain amount for each ton sold abroad. How does this export subsidy affect the domestic price of steel, the quantity of steel produced, the quantity of steel consumed, and the quantity of steel exported? How does it affect consumer surplus, producer surplus, government revenue, and total surplus? Is it a good policy from the standpoint of economic efficiency? (Hint: The analysis of an export subsidy is similar to the analysis of a tariff.)
This page intentionally left blank
PART IV

The Economics of the Public Sector
Firms that make and sell paper also create, as a by-product of the manufacturing process, a chemical called dioxin. Scientists believe that once dioxin enters the environment, it raises the population’s risk of cancer, birth defects, and other health problems.

Is the production and release of dioxin a problem for society? In Chapters 4 through 9, we examined how markets allocate scarce resources with the forces of supply and demand, and we saw that the equilibrium of supply and demand is typically an efficient allocation of resources. To use Adam Smith’s famous metaphor, the “invisible hand” of the marketplace leads self-interested buyers and sellers in a market to maximize the total benefit that society derives from that market. This insight is the basis for one of the Ten Principles of Economics in Chapter 1: Markets are usually a good way to organize economic activity. Should we conclude, therefore, that the invisible hand prevents firms in the paper market from emitting too much dioxin?

Markets do many things well, but they do not do everything well. In this chapter, we begin our study of another of the Ten Principles of Economics: Government action can sometimes improve upon market outcomes. We examine why markets sometimes fail to allocate resources efficiently, how government policies can potentially improve the market’s allocation, and what kinds of policies are likely to work best.

Externalities
The market failures examined in this chapter fall under a general category called *externalities*. An *externality* arises when a person engages in an activity that influences the well-being of a bystander and yet neither pays nor receives any compensation for that effect. If the impact on the bystander is adverse, it is called a *negative externality*. If it is beneficial, it is called a *positive externality*. In the presence of externalities, society’s interest in a market outcome extends beyond the well-being of buyers and sellers who participate in the market to include the well-being of bystanders who are affected indirectly. Because buyers and sellers neglect the external effects of their actions when deciding how much to demand or supply, the market equilibrium is not efficient when there are externalities. That is, the equilibrium fails to maximize the total benefit to society as a whole. The release of dioxin into the environment, for instance, is a negative externality. Self-interested paper firms will not consider the full cost of the pollution they create in their production process, and consumers of paper will not consider the full cost of the pollution they contribute from their purchasing decisions. Therefore, the firms will emit too much pollution unless the government prevents or discourages them from doing so.

Externalities come in many varieties, as do the policy responses that try to deal with the market failure. Here are some examples:

- The exhaust from automobiles is a negative externality because it creates smog that other people have to breathe. As a result of this externality, drivers tend to pollute too much. The federal government attempts to solve this problem by setting emission standards for cars. It also taxes gasoline to reduce the amount that people drive.

- Restored historic buildings convey a positive externality because people who walk or ride by them can enjoy the beauty and the sense of history that these buildings provide. Building owners do not get the full benefit of restoration and, therefore, tend to discard older buildings too quickly. Many local governments respond to this problem by regulating the destruction of historic buildings and by providing tax breaks to owners who restore them.

- Barking dogs create a negative externality because neighbors are disturbed by the noise. Dog owners do not bear the full cost of the noise and, therefore, tend to take too few precautions to prevent their dogs from barking. Local governments address this problem by making it illegal to “disturb the peace.”

- Research into new technologies provides a positive externality because it creates knowledge that other people can use. Because inventors cannot capture the full benefits of their inventions, they tend to devote too few resources to research. The federal government addresses this problem partially through the patent system, which gives inventors exclusive use of their inventions for a limited time.

In each of these cases, some decision maker fails to take account of the external effects of his or her behavior. The government responds by trying to influence this behavior to protect the interests of bystanders.

**EXTERNALITIES AND MARKET INEFFICIENCY**

In this section, we use the tools of welfare economics developed in Chapter 7 to examine how externalities affect economic well-being. The analysis shows precisely why externalities cause markets to allocate resources inefficiently. Later in
the chapter, we examine various ways in which private individuals and public policymakers may remedy this type of market failure.

**Welfare Economics: A Recap**

We begin by recalling the key lessons of welfare economics from Chapter 7. To make our analysis concrete, we consider a specific market—the market for aluminum. Figure 1 shows the supply and demand curves in the market for aluminum.

As you should recall from Chapter 7, the supply and demand curves contain important information about costs and benefits. The demand curve for aluminum reflects the value of aluminum to consumers, as measured by the prices they are willing to pay. At any given quantity, the height of the demand curve shows the willingness to pay of the marginal buyer. In other words, it shows the value to the consumer of the last unit of aluminum bought. Similarly, the supply curve reflects the costs of producing aluminum. At any given quantity, the height of the supply curve shows the cost of the marginal seller. In other words, it shows the cost to the producer of the last unit of aluminum sold.

In the absence of government intervention, the price adjusts to balance the supply and demand for aluminum. The quantity produced and consumed in the market equilibrium, shown as $Q_{\text{MARKET}}$ in Figure 1, is efficient in the sense that it maximizes the sum of producer and consumer surplus. That is, the market allocates resources in a way that maximizes the total value to the consumers who buy and use aluminum minus the total costs to the producers who make and sell aluminum.

**Negative Externalities**

Now let’s suppose that aluminum factories emit pollution: For each unit of aluminum produced, a certain amount of smoke enters the atmosphere. Because this
smoke creates a health risk for those who breathe the air, it is a negative externality. How does this externality affect the efficiency of the market outcome?

Because of the externality, the cost to society of producing aluminum is larger than the cost to the aluminum producers. For each unit of aluminum produced, the social cost includes the private costs of the aluminum producers plus the costs to those bystanders affected adversely by the pollution. Figure 2 shows the social cost of producing aluminum. The social-cost curve is above the supply curve because it takes into account the external costs imposed on society by aluminum producers. The difference between these two curves reflects the cost of the pollution emitted.

What quantity of aluminum should be produced? To answer this question, we once again consider what a benevolent social planner would do. The planner wants to maximize the total surplus derived from the market—the value to consumers of aluminum minus the cost of producing aluminum. The planner understands, however, that the cost of producing aluminum includes the external costs of the pollution.

The planner would choose the level of aluminum production at which the demand curve crosses the social-cost curve. This intersection determines the optimal amount of aluminum from the standpoint of society as a whole. Below this level of production, the value of the aluminum to consumers (as measured by the height of the demand curve) exceeds the social cost of producing it (as measured by the height of the social-cost curve). The planner does not produce more than this level because the social cost of producing additional aluminum exceeds the value to consumers.

Note that the equilibrium quantity of aluminum, \( Q_{\text{MARKET}} \), is larger than the socially optimal quantity, \( Q_{\text{OPTIMUM}} \). This inefficiency occurs because the market equilibrium reflects only the private costs of production. In the market equilib-
rium, the marginal consumer values aluminum at less than the social cost of producing it. That is, at $Q_{\text{MARKET}}$, the demand curve lies below the social-cost curve. Thus, reducing aluminum production and consumption below the market equilibrium level raises total economic well-being.

How can the social planner achieve the optimal outcome? One way would be to tax aluminum producers for each ton of aluminum sold. The tax would shift the supply curve for aluminum upward by the size of the tax. If the tax accurately reflected the external cost of smoke released into the atmosphere, the new supply curve would coincide with the social-cost curve. In the new market equilibrium, aluminum producers would produce the socially optimal quantity of aluminum.

The use of such a tax is called **internalizing the externality** because it gives buyers and sellers in the market an incentive to take into account the external effects of their actions. Aluminum producers would, in essence, take the costs of pollution into account when deciding how much aluminum to supply because the tax would make them pay for these external costs. And, because the market price would reflect the tax on producers, consumers of aluminum would have an incentive to use a smaller quantity. The policy is based on one of the **Ten Principles of Economics**: People respond to incentives. Later in this chapter, we consider in more detail how policymakers can deal with externalities.

### Positive Externalities

Although some activities impose costs on third parties, others yield benefits. For example, consider education. To a large extent, the benefit of education is private: The consumer of education becomes a more productive worker and thus reaps much of the benefit in the form of higher wages. Beyond these private benefits, however, education also yields positive externalities. One externality is that a more educated population leads to more informed voters, which means better government for everyone. Another externality is that a more educated population tends to mean lower crime rates. A third externality is that a more educated population may encourage the development and dissemination of technological advances, leading to higher productivity and wages for everyone. Because of these three positive externalities, a person may prefer to have neighbors who are well educated.

The analysis of positive externalities is similar to the analysis of negative externalities. As Figure 3 shows, the demand curve does not reflect the value to society of the good. Because the social value is greater than the private value, the social-value curve lies above the demand curve. The optimal quantity is found where the social-value curve and the supply curve (which represents costs) intersect. Hence, the socially optimal quantity is greater than the quantity determined by the private market.

Once again, the government can correct the market failure by inducing market participants to internalize the externality. The appropriate response in the case of positive externalities is exactly the opposite to the case of negative externalities. To move the market equilibrium closer to the social optimum, a positive externality requires a subsidy. In fact, that is exactly the policy the government follows: Education is heavily subsidized through public schools and government scholarships.

To summarize: **Negative externalities lead markets to produce a larger quantity than is socially desirable. Positive externalities lead markets to produce a smaller quantity than**
is socially desirable. To remedy the problem, the government can internalize the externality by taxing goods that have negative externalities and subsidizing goods that have positive externalities.

**Technology Spillovers, Industrial Policy, and Patent Protection**

A potentially important type of positive externality is called a *technology spillover*—the impact of one firm’s research and production efforts on other firms’ access to technological advance. For example, consider the market for industrial robots. Robots are at the frontier of a rapidly changing technology. Whenever a firm builds a robot, there is some chance that it will discover a new and better design. This new design may benefit not only this firm but society as a whole because the design will enter society’s pool of technological knowledge. That is, the new design may have positive externalities for other producers in the economy.

In this case, the government can internalize the externality by subsidizing the production of robots. If the government paid firms a subsidy for each robot produced, the supply curve would shift down by the amount of the subsidy, and this shift would increase the equilibrium quantity of robots. To ensure that the market equilibrium equals the social optimum, the subsidy should equal the value of the technology spillover.

How large are technology spillovers, and what do they imply for public policy? This is an important question because technological progress is the key to why living standards rise over time. Yet it is also a difficult question on which economists often disagree.

Some economists believe that technology spillovers are pervasive and that the government should encourage those industries that yield the largest spillovers.
For instance, these economists argue that if making computer chips yields greater spillovers than making potato chips, then the government should encourage the production of computer chips relative to the production of potato chips. The U.S. tax code does this in a limited way by offering special tax breaks for expenditures on research and development. Some other nations go further by subsidizing specific industries that supposedly offer large technology spillovers. Government intervention in the economy that aims to promote technology-enhancing industries is sometimes called industrial policy.

Other economists are skeptical about industrial policy. Even if technology spillovers are common, the success of an industrial policy requires that the government be able to measure the size of the spillovers from different markets. This measurement problem is difficult at best. Moreover, without precise measurements, the political system may end up subsidizing industries with the most political clout rather than those that yield the largest positive externalities.

Another way to deal with technology spillovers is patent protection. The patent laws protect the rights of inventors by giving them exclusive use of their inventions for a period of time. When a firm makes a technological breakthrough, it can patent the idea and capture much of the economic benefit for itself. The patent internalizes the externality by giving the firm a property right over its invention. If other firms want to use the new technology, they would have to obtain permission from the inventing firm and pay it some royalty. Thus, the patent system gives firms a greater incentive to engage in research and other activities that advance technology.

Quick Quiz  Give an example of a negative externality and a positive externality. Explain why market outcomes are inefficient in the presence of these externalities.

Public Policies Toward Externalities

We have discussed why externalities lead markets to allocate resources inefficiently but have mentioned only briefly how this inefficiency can be remedied. In practice, both public policymakers and private individuals respond to externalities in various ways. All of the remedies share the goal of moving the allocation of resources closer to the social optimum.

This section considers governmental solutions. As a general matter, the government can respond to externalities in one of two ways. Command-and-control policies regulate behavior directly. Market-based policies provide incentives so that private decision makers will choose to solve the problem on their own.

Command-and-Control Policies: Regulation

The government can remedy an externality by making certain behaviors either required or forbidden. For example, it is a crime to dump poisonous chemicals into the water supply. In this case, the external costs to society far exceed the benefits to the polluter. The government therefore institutes a command-and-control policy that prohibits this act altogether.

In most cases of pollution, however, the situation is not this simple. Despite the stated goals of some environmentalists, it would be impossible to prohibit all polluting activity. For example, virtually all forms of transportation—even
the horse—produce some undesirable polluting by-products. But it would not be sensible for the government to ban all transportation. Thus, instead of trying to eradicate pollution entirely, society has to weigh the costs and benefits to decide the kinds and quantities of pollution it will allow. In the United States, the Environmental Protection Agency (EPA) is the government agency with the task of developing and enforcing regulations aimed at protecting the environment.

Environmental regulations can take many forms. Sometimes the EPA dictates a maximum level of pollution that a factory may emit. Other times the EPA requires that firms adopt a particular technology to reduce emissions. In all cases, to design good rules, the government regulators need to know the details about specific industries and about the alternative technologies that those industries could adopt. This information is often difficult for government regulators to obtain.

**MARKET-BASED POLICY 1: CORRECTIVE TAXES AND SUBSIDIES**

Instead of regulating behavior in response to an externality, the government can use market-based policies to align private incentives with social efficiency. For instance, as we saw earlier, the government can internalize the externality by taxing activities that have negative externalities and subsidizing activities that have positive externalities. Taxes enacted to deal with the effects of negative externalities are called **corrective taxes**. They are also called *Pigovian taxes* after economist Arthur Pigou (1877–1959), an early advocate of their use. An ideal corrective tax would equal the external cost from an activity with negative externalities, and an ideal corrective subsidy would equal the external benefit from an activity with positive externalities.

Economists usually prefer corrective taxes to regulations as a way to deal with pollution because they can reduce pollution at a lower cost to society. To see why, let us consider an example.

Suppose that two factories—a paper mill and a steel mill—are each dumping 500 tons of glop into a river each year. The EPA decides that it wants to reduce the amount of pollution. It considers two solutions:

- **Regulation**: The EPA could tell each factory to reduce its pollution to 300 tons of glop per year.
- **Corrective tax**: The EPA could levy a tax on each factory of $50,000 for each ton of glop it emits.

The regulation would dictate a level of pollution, whereas the tax would give factory owners an economic incentive to reduce pollution. Which solution do you think is better?

Most economists prefer the tax. To explain this preference, they would first point out that a tax is just as effective as a regulation in reducing the overall level of pollution. The EPA can achieve whatever level of pollution it wants by setting the tax at the appropriate level. The higher the tax, the larger the reduction in pollution. If the tax is high enough, the factories will close down altogether, reducing pollution to zero.

Although regulation and corrective taxes are both capable of reducing pollution, the tax accomplishes this goal more efficiently. The regulation requires each factory to reduce pollution by the same amount. An equal reduction, however, is not necessarily the least expensive way to clean up the water. It is possible that the paper
mill can reduce pollution at lower cost than the steel mill. If so, the paper mill would respond to the tax by reducing pollution substantially to avoid the tax, whereas the steel mill would respond by reducing pollution less and paying the tax.

In essence, the corrective tax places a price on the right to pollute. Just as markets allocate goods to those buyers who value them most highly, a corrective tax allocates pollution to those factories that face the highest cost of reducing it. Whatever the level of pollution the EPA chooses, it can achieve this goal at the lowest total cost using a tax.

Economists also argue that corrective taxes are better for the environment. Under the command-and-control policy of regulation, the factories have no reason to reduce emission further once they have reached the target of 300 tons of glop. By contrast, the tax gives the factories an incentive to develop cleaner technologies because a cleaner technology would reduce the amount of tax the factory has to pay.

Corrective taxes are unlike most other taxes. As we discussed in Chapter 8, most taxes distort incentives and move the allocation of resources away from the social optimum. The reduction in economic well-being—that is, in consumer and producer surplus—exceeds the amount of revenue the government raises, resulting in a deadweight loss. By contrast, when externalities are present, society also cares about the well-being of the bystanders who are affected. Corrective taxes alter incentives to account for the presence of externalities and thereby move the allocation of resources closer to the social optimum. Thus, while corrective taxes raise revenue for the government, they also enhance economic efficiency.

**WHY IS GASOLINE TAXED SO HEAVILY?**

In many nations, gasoline is among the most heavily taxed goods. The gas tax can be viewed as a corrective tax aimed at three negative externalities associated with driving:

- **Congestion:** If you have ever been stuck in bumper-to-bumper traffic, you have probably wished that there were fewer cars on the road. A gasoline tax keeps congestion down by encouraging people to take public transportation, carpool more often, and live closer to work.
- **Accidents:** Whenever people buy large cars or sport-utility vehicles, they may make themselves safer but they certainly put their neighbors at risk. According to the National Highway Traffic Safety Administration, a person driving a typical car is five times as likely to die if hit by a sport-utility vehicle than if hit by another car. The gas tax is an indirect way of making people pay when their large, gas-guzzling vehicles impose risk on others, which in turn makes them take this risk into account when choosing what vehicle to purchase.
- **Pollution:** The burning of fossil fuels such as gasoline is widely believed to be the cause of global warming. Experts disagree about how dangerous this threat is, but there is no doubt that the gas tax reduces the threat by reducing the use of gasoline.

So the gas tax, rather than causing deadweight losses like most taxes, actually makes the economy work better. It means less traffic congestion, safer roads, and a cleaner environment.
How high should the tax on gasoline be? Most European countries impose gasoline taxes that are much higher than those in the United States. Many observers have suggested that the United States also should tax gasoline more heavily. A 2007 study published in the *Journal of Economic Literature* summarized the research on the size of the various externalities associated with driving. It concluded that the optimal corrective tax on gasoline was $2.10 per gallon, compared to the actual tax in the United States of 40 cents.

The tax revenue from a gasoline tax could be used to lower taxes that distort incentives and cause deadweight losses. In addition, some of the burdensome government regulations that require automakers to produce more fuel-efficient cars would prove unnecessary. This idea, however, has never proven politically popular.

**Market-Based Policy 2: Tradable Pollution Permits**

Returning to our example of the paper mill and the steel mill, let us suppose that, despite the advice of its economists, the EPA adopts the regulation and requires each factory to reduce its pollution to 300 tons of glop per year. Then one day, after the regulation is in place and both mills have complied, the two firms go to the EPA with a proposal. The steel mill wants to increase its emission of glop by 100 tons. The paper mill has agreed to reduce its emission by the same amount if the steel mill pays it $5 million. Should the EPA allow the two factories to make this deal?

From the standpoint of economic efficiency, allowing the deal is good policy. The deal must make the owners of the two factories better off because they are
voluntarily agreeing to it. Moreover, the deal does not have any external effects because the total amount of pollution remains the same. Thus, social welfare is enhanced by allowing the paper mill to sell its pollution rights to the steel mill.

The same logic applies to any voluntary transfer of the right to pollute from one firm to another. If the EPA allows firms to make these deals, it will, in essence, have created a new scarce resource: pollution permits. A market to trade these permits will eventually develop, and that market will be governed by the forces of supply and demand. The invisible hand will ensure that this new market allocates the right to pollute efficiently. That is, the permits will end up in the hands of those firms that value them most highly, as judged by their willingness to pay. A firm’s willingness to pay, in turn, will depend on its cost of reducing pollution: The more costly it is for a firm to cut back on pollution, the more it will be willing to pay for a permit.

An advantage of allowing a market for pollution permits is that the initial allocation of pollution permits among firms does not matter from the standpoint of economic efficiency. Those firms that can reduce pollution at a low cost will sell whatever permits they get, and firms that can reduce pollution only at a high cost will buy whatever permits they need. As long as there is a free market for the pollution rights, the final allocation will be efficient regardless of the initial allocation.

Although reducing pollution using pollution permits may seem very different from using corrective taxes, the two policies have much in common. In both cases, firms pay for their pollution. With corrective taxes, polluting firms must pay a tax to the government. With pollution permits, polluting firms must pay to buy the permit. (Even firms that already own permits must pay to pollute: The opportunity cost of polluting is what they could have received by selling their permits on the open market.) Both corrective taxes and pollution permits internalize the externality of pollution by making it costly for firms to pollute.

The similarity of the two policies can be seen by considering the market for pollution. Both panels in Figure 4 show the demand curve for the right to pollute. This curve shows that the lower the price of polluting, the more firms will choose to pollute. In panel (a), the EPA uses a corrective tax to set a price for pollution. In this case, the supply curve for pollution rights is perfectly elastic (because firms can pollute as much as they want by paying the tax), and the position of the demand curve determines the quantity of pollution. In panel (b), the EPA sets a quantity of pollution by issuing pollution permits. In this case, the supply curve for pollution rights is perfectly inelastic (because the quantity of pollution is fixed by the number of permits), and the position of the demand curve determines the price of pollution. Hence, the EPA can achieve any point on a given demand curve either by setting a price with a corrective tax or by setting a quantity with pollution permits.

In some circumstances, however, selling pollution permits may be better than levying a corrective tax. Suppose the EPA wants no more than 600 tons of gloop dumped into the river. But because the EPA does not know the demand curve for pollution, it is not sure what size tax would achieve that goal. In this case, it can simply auction off 600 pollution permits. The auction price would yield the appropriate size of the corrective tax.

The idea of the government auctioning off the right to pollute may at first sound like a creature of some economist’s imagination. And in fact, that is how the idea began. But increasingly, the EPA has used the system as a way to control pollution. A notable success story has been the case of sulfur dioxide (SO₂)—a
leading cause of acid rain. In 1990, amendments to the Clean Air Act required power plants to reduce SO₂ emissions substantially. At the same time, the amendments set up a system that allowed plants to trade their SO₂ allowances. Although initially both industry representatives and environmentalists were skeptical of the proposal, over time the system has proved that it can reduce pollution with minimal disruption. Pollution permits, like corrective taxes, are now widely viewed as a cost-effective way to keep the environment clean.

**Objections to the Economic Analysis of Pollution**

“We cannot give anyone the option of polluting for a fee.” This comment by former Senator Edmund Muskie reflects the view of some environmentalists. Clean air and clean water, they argue, are fundamental human rights that should not be debased by considering them in economic terms. How can you put a price on clean air and clean water? The environment is so important, they claim, that we should protect it as much as possible, regardless of the cost.

Economists have little sympathy for this type of argument. To economists, good environmental policy begins by acknowledging the first of the *Ten Principles of Economics* in Chapter 1: People face trade-offs. Certainly, clean air and clean water have value. But their value must be compared to their opportunity cost—that is, to what one must give up to obtain them. Eliminating all pollution is impossible. Trying to eliminate all pollution would reverse many of the technological advances that allow us to enjoy a high standard of living. Few people would be
willing to accept poor nutrition, inadequate medical care, or shoddy housing to make the environment as clean as possible.

Economists argue that some environmental activists hurt their own cause by not thinking in economic terms. A clean environment is a good like other goods. Like all normal goods, it has a positive income elasticity: Rich countries can afford a cleaner environment than poor ones and, therefore, usually have more rigorous environmental protection. In addition, like most other goods, clean air and clean water obey the law of demand: The lower the price of environmental protection, the more the public will want. The economic approach of using pollution permits and corrective taxes reduces the cost of environmental protection and should, therefore, increase the public’s demand for a clean environment.

**Quick Quiz**  A glue factory and a steel mill emit smoke containing a chemical that is harmful if inhaled in large amounts. Describe three ways the town government might respond to this externality. What are the pros and cons of each solution?

---

**PRIVATE SOLUTIONS TO EXTERNALITIES**

Although externalities tend to cause markets to be inefficient, government action is not always needed to solve the problem. In some circumstances, people can develop private solutions.

**The Types of Private Solutions**

Sometimes the problem of externalities is solved with moral codes and social sanctions. Consider, for instance, why most people do not litter. Although there are laws against littering, these laws are not vigorously enforced. Most people do not litter just because it is the wrong thing to do. The Golden Rule taught to most children says, “Do unto others as you would have them do unto you.” This moral injunction tells us to take account of how our actions affect other people. In economic terms, it tells us to internalize externalities.

Another private solution to externalities is charities, many of which are established to deal with externalities. For example, the Sierra Club, whose goal is to protect the environment, is a nonprofit organization funded with private donations. As another example, colleges and universities receive gifts from alumni, corporations, and foundations in part because education has positive externalities for society. The government encourages this private solution to externalities through the tax system by allowing an income tax deduction for charitable donations.

The private market can often solve the problem of externalities by relying on the self-interest of the relevant parties. Sometimes the solution takes the form of integrating different types of businesses. For example, consider an apple grower and a beekeeper who are located next to each other. Each business confers a positive externality on the other: By pollinating the flowers on the trees, the bees help the orchard produce apples. At the same time, the bees use the nectar they get from the apple trees to produce honey. Nonetheless, when the apple grower is deciding how many trees to plant and the beekeeper is deciding how many bees to keep, they neglect the positive externality. As a result, the apple grower plants too few trees and the beekeeper keeps too few bees. These externalities could be internalized if the beekeeper bought the apple orchard or if the apple grower...
bought the beehives: Both activities would then take place within the same firm, and this single firm could choose the optimal number of trees and bees. Internalizing externalities is one reason that some firms are involved in different types of businesses.

Another way for the private market to deal with external effects is for the interested parties to enter into a contract. In the foregoing example, a contract between the apple grower and the beekeeper can solve the problem of too few trees and too few bees. The contract can specify the number of trees, the number of bees, and perhaps a payment from one party to the other. By setting the right number of trees and bees, the contract can solve the inefficiency that normally arises from these externalities and make both parties better off.

**One Answer to Global Warming: A New Tax**

*By N. Gregory Mankiw*

In the debate over global climate change, there is a yawning gap that needs to be bridged. The gap is not between environmentalists and industrialists, or between Democrats and Republicans. It is between policy wonks and political consultants.

Among policy wonks like me, there is a broad consensus. The scientists tell us that world temperatures are rising because humans are emitting carbon into the atmosphere. Basic economics tells us that when you tax something, you normally get less of it. So if we want to reduce global emissions of carbon, we need a global carbon tax. QED.

The idea of using taxes to fix problems, rather than merely raise government revenue, has a long history. The British economist Arthur Pigou advocated such corrective taxes to deal with pollution in the early 20th century. In his honor, economics textbooks now call them “Pigovian taxes.”

Using a Pigovian tax to address global warming is also an old idea. It was proposed as far back as 1992 by Martin S. Feldstein on the editorial page of *The Wall Street Journal*. Once chief economist to Ronald Reagan, Mr. Feldstein has devoted much of his career to studying how high tax rates distort incentives and impede economic growth. But like most other policy wonks, he appreciates that some taxes align private incentives with social costs and move us toward better outcomes.

Those vying for elected office, however, are reluctant to sign on to this agenda. Their political consultants are no fans of taxes, Pigovian or otherwise. Republican consultants advise using the word “tax” only if followed immediately by the word “cut.” Democratic consultants recommend the word “tax” be followed by “on the rich.”

Yet this natural aversion to carbon taxes can be overcome if the revenue from the tax is used to reduce other taxes. By itself, a carbon tax would raise the tax burden on anyone who drives a car or uses electricity produced with fossil fuels, which means just about everybody. Some might fear this would be particularly hard on the poor and middle class.

But Gilbert Metcalf, a professor of economics at Tufts, has shown how revenue from a carbon tax could be used to reduce payroll taxes in a way that would leave the distribution of total tax burden approximately unchanged. He proposes a tax of $15 per metric ton of carbon dioxide, together with a rebate of the federal payroll tax on the first $3,660 of earnings for each worker.

The case for a carbon tax looks even stronger after an examination of the other options on the table. Lawmakers in both political parties want to require carmakers to increase the fuel efficiency of the cars they sell. Passing the buck to auto companies has a lot of popular appeal.

Increased fuel efficiency, however, is not free. Like a tax, the cost of complying with...
The Coase Theorem

How effective is the private market in dealing with externalities? A famous result, called the Coase theorem after economist Ronald Coase, suggests that it can be very effective in some circumstances. According to the Coase theorem, if private parties can bargain over the allocation of resources at no cost, then the private market will always solve the problem of externalities and allocate resources efficiently.

To see how the Coase theorem works, consider an example. Suppose that Dick owns a dog named Spot. Spot barks and disturbs Jane, Dick’s neighbor. Dick gets a benefit from owning the dog, but the dog confers a negative externality on Jane. More stringent regulation will be passed on to consumers in the form of higher car prices. But the government will not raise any revenue that it can use to cut other taxes to compensate for these higher prices. (And don’t expect savings on gas to compensate consumers in a meaningful way: Any truly cost-effective increase in fuel efficiency would already have been made.)

More important, enhancing fuel efficiency by itself is not the best way to reduce energy consumption. Fuel use depends not only on the efficiency of the car fleet but also on the daily decisions that people make—how far from work they choose to live and how often they carpool or use public transportation.

A carbon tax would provide incentives for people to use less fuel in a multitude of ways. By contrast, merely having more efficient cars encourages more driving. Increased driving not only produces more carbon but also exacerbates other problems, like accidents and road congestion.

Another popular proposal to limit carbon emissions is a cap-and-trade system, under which carbon emissions are limited and allowances are bought and sold in the marketplace. The effect of such a system depends on how the carbon allowances are allocated. If the government auctions them off, then the price of a carbon allowance is effectively a carbon tax.

But the history of cap-and-trade systems suggests that the allowances would probably be handed out to power companies and other carbon emitters, which would then be free to use them or sell them at market prices. In this case, the prices of energy products would rise as they would under a carbon tax, but the government would collect no revenue to reduce other taxes and compensate consumers.

The international dimension of the problem also suggests the superiority of a carbon tax over cap-and-trade. Any long-term approach to global climate change will have to deal with the emerging economies of China and India. By some reports, China is now the world’s leading emitter of carbon, in large part simply because it has so many people. The failure of the Kyoto treaty to include these emerging economies is one reason that, in 1997, the United States Senate passed a resolution rejecting the Kyoto approach by a vote of 95 to zero.

Agreement on a truly global cap-and-trade system, however, is hard to imagine. China is unlikely to be persuaded to accept fewer carbon allowances per person than the United States. Using a historical baseline to allocate allowances, as is often proposed, would reward the United States for having been a leading cause of the problem. But allocating carbon allowances based on population alone would create a system in which the United States, with its higher standard of living, would buy allowances from China. American voters are not going to embrace a system of higher energy prices, coupled with a large transfer of national income to the Chinese. It would amount to a massive foreign aid program to one of the world’s most rapidly growing economies.

A global carbon tax would be easier to negotiate. All governments require revenue for public purposes. The world’s nations could agree to use a carbon tax as one instrument to raise some of that revenue. No money needs to change hands across national borders. Each government could keep the revenue from its tax and use it to finance spending or whatever form of tax relief it considered best. Convincing China of the virtues of a carbon tax, however, may prove to be the easy part. The first and more difficult step is to convince American voters, and therefore political consultants, that “tax” is not a four-letter word.

Should Dick be forced to send Spot to the pound, or should Jane have to suffer sleepless nights because of Spot’s barking?

Consider first what outcome is socially efficient. A social planner, considering the two alternatives, would compare the benefit that Dick gets from the dog to the cost that Jane bears from the barking. If the benefit exceeds the cost, it is efficient for Dick to keep the dog and for Jane to live with the barking. Yet if the cost exceeds the benefit, then Dick should get rid of the dog.

According to the Coase theorem, the private market will reach the efficient outcome on its own. How? Jane can simply offer to pay Dick to get rid of the dog. Dick will accept the deal if the amount of money Jane offers is greater than the benefit of keeping the dog.

By bargaining over the price, Dick and Jane can always reach the efficient outcome. For instance, suppose that Dick gets a $500 benefit from the dog and Jane bears an $800 cost from the barking. In this case, Jane can offer Dick $600 to get rid of the dog, and Dick will gladly accept. Both parties are better off than they were before, and the efficient outcome is reached.

It is possible, of course, that Jane would not be willing to offer any price that Dick would accept. For instance, suppose that Dick gets a $1,000 benefit from the dog and Jane bears an $800 cost from the barking. In this case, Dick would turn down any offer below $1,000, while Jane would not offer any amount above $800. Therefore, Dick ends up keeping the dog. Given these costs and benefits, however, this outcome is efficient.

So far, we have assumed that Dick has the legal right to keep a barking dog. In other words, we have assumed that Dick can keep Spot unless Jane pays him enough to induce him to give up the dog voluntarily. But how different would the outcome be if Jane had the legal right to peace and quiet?

According to the Coase theorem, the initial distribution of rights does not matter for the market’s ability to reach the efficient outcome. For instance, suppose that Jane can legally compel Dick to get rid of the dog. Although having this right works to Jane’s advantage, it probably will not change the outcome. In this case, Dick can offer to pay Jane to allow him to keep the dog. If the benefit of the dog to Dick exceeds the cost of the barking to Jane, then Dick and Jane will strike a bargain in which Dick keeps the dog.

Although Dick and Jane can reach the efficient outcome regardless of how rights are initially distributed, the distribution of rights is not irrelevant: It determines the distribution of economic well-being. Whether Dick has the right to a barking dog or Jane the right to peace and quiet determines who pays whom in the final bargain. But in either case, the two parties can bargain with each other and solve the externality problem. Dick will end up keeping the dog only if the benefit exceeds the cost.

To sum up: The Coase theorem says that private economic actors can solve the problem of externalities among themselves. Whatever the initial distribution of rights, the interested parties can always reach a bargain in which everyone is better off and the outcome is efficient.

**Why Private Solutions Do Not Always Work**

Despite the appealing logic of the Coase theorem, private individuals on their own often fail to resolve the problems caused by externalities. The Coase theorem applies only when the interested parties have no trouble reaching and enforcing
an agreement. In the real world, however, bargaining does not always work, even when a mutually beneficial agreement is possible.

Sometimes the interested parties fail to solve an externality problem because of transaction costs, the costs that parties incur in the process of agreeing to and following through on a bargain. In our example, imagine that Dick and Jane speak different languages so that, to reach an agreement, they need to hire a translator. If the benefit of solving the barking problem is less than the cost of the translator, Dick and Jane might choose to leave the problem unsolved. In more realistic examples, the transaction costs are the expenses not of translators but of the lawyers required to draft and enforce contracts.

At other times, bargaining simply breaks down. The recurrence of wars and labor strikes shows that reaching agreement can be difficult and that failing to reach agreement can be costly. The problem is often that each party tries to hold out for a better deal. For example, suppose that Dick gets a $500 benefit from the dog, and Jane bears an $800 cost from the barking. Although it is efficient for Jane to pay Dick to get rid of the dog, there are many prices that could lead to this outcome. Dick might demand $750, and Jane might offer only $550. As they haggle over the price, the inefficient outcome with the barking dog persists.

Reaching an efficient bargain is especially difficult when the number of interested parties is large because coordinating everyone is costly. For example, consider a factory that pollutes the water of a nearby lake. The pollution confers a negative externality on the local fishermen. According to the Coase theorem, if the pollution is inefficient, then the factory and the fishermen could reach a bargain in which the fishermen pay the factory not to pollute. If there are many fishermen, however, trying to coordinate them all to bargain with the factory may be almost impossible.

When private bargaining does not work, the government can sometimes play a role. The government is an institution designed for collective action. In this example, the government can act on behalf of the fishermen, even when it is impractical for the fishermen to act for themselves.

**Quick Quiz**  Give an example of a private solution to an externality. • What is the Coase theorem? • Why are private economic participants sometimes unable to solve the problems caused by an externality?

**Conclusion**

The invisible hand is powerful but not omnipotent. A market’s equilibrium maximizes the sum of producer and consumer surplus. When the buyers and sellers in the market are the only interested parties, this outcome is efficient from the standpoint of society as a whole. But when there are external effects, such as pollution, evaluating a market outcome requires taking into account the well-being of third parties as well. In this case, the invisible hand of the marketplace may fail to allocate resources efficiently.

In some cases, people can solve the problem of externalities on their own. The Coase theorem suggests that the interested parties can bargain among themselves and agree on an efficient solution. Sometimes, however, an efficient outcome cannot be reached, perhaps because the large number of interested parties makes bargaining difficult.
When people cannot solve the problem of externalities privately, the government often steps in. Yet even with government intervention, society should not abandon market forces entirely. Rather, the government can address the problem by requiring decision makers to bear the full costs of their actions. Corrective taxes on emissions and pollution permits, for instance, are designed to internalize the externality of pollution. More and more, these are the policies of choice for those interested in protecting the environment. Market forces, properly redirected, are often the best remedy for market failure.

**SUMMARY**

- When a transaction between a buyer and seller directly affects a third party, the effect is called an externality. If an activity yields negative externalities, such as pollution, the socially optimal quantity in a market is less than the equilibrium quantity. If an activity yields positive externalities, such as technology spillovers, the socially optimal quantity is greater than the equilibrium quantity.

- Governments pursue various policies to remedy the inefficiencies caused by externalities. Sometimes the government prevents socially inefficient activity by regulating behavior. Other times it internalizes an externality using corrective taxes. Another public policy is to issue permits. For example, the government could protect the environment by issuing a limited number of pollution permits. The result of this policy is largely the same as imposing corrective taxes on polluters.

- Those affected by externalities can sometimes solve the problem privately. For instance, when one business imposes an externality on another business, the two businesses can internalize the externality by merging. Alternatively, the interested parties can solve the problem by negotiating a contract. According to the Coase theorem, if people can bargain without cost, then they can always reach an agreement in which resources are allocated efficiently. In many cases, however, reaching a bargain among the many interested parties is difficult, so the Coase theorem does not apply.

**KEY CONCEPTS**

- externality, p. 204
- internalizing the externality, p. 207
- corrective tax, p. 210
- Coase theorem, p. 217
- transaction costs, p. 219
QUESTIONS FOR REVIEW

1. Give an example of a negative externality and an example of a positive externality.
2. Draw a supply-and-demand diagram to explain the effect of a negative externality that occurs as a result of a firm’s production process.
3. In what way does the patent system help society solve an externality problem?
4. What are corrective taxes? Why do economists prefer them to regulations as a way to protect the environment from pollution?
5. List some of the ways that the problems caused by externalities can be solved without government intervention.
6. Imagine that you are a nonsmoker sharing a room with a smoker. According to the Coase theorem, what determines whether your roommate smokes in the room? Is this outcome efficient? How do you and your roommate reach this solution?

PROBLEMS AND APPLICATIONS

1. There are two ways to protect your car from theft. The Club makes it difficult for a car thief to take your car. Lojack makes it easier for the police to catch the car thief who has stolen it. Which of these types of protection conveys a negative externality on other car owners? Which conveys a positive externality? Do you think there are any policy implications of your analysis?
2. Do you agree with the following statements? Why or why not?
   a. “The benefits of corrective taxes as a way to reduce pollution have to be weighed against the deadweight losses that these taxes cause.”
   b. “When deciding whether to levy a corrective tax on consumers or producers, the government should be careful to levy the tax on the side of the market generating the externality.”
3. Consider the market for fire extinguishers.
   a. Why might fire extinguishers exhibit positive externalities?
   b. Draw a graph of the market for fire extinguishers, labeling the demand curve, the social-value curve, the supply curve, and the social-cost curve.
   c. Indicate the market equilibrium level of output and the efficient level of output. Give an intuitive explanation for why these quantities differ.
   d. If the external benefit is $10 per extinguisher, describe a government policy that would yield the efficient outcome.
4. It is rumored that the Swiss government subsidizes cattle farming and that the subsidy is larger in areas with more tourist attractions. Can you think of a reason this policy might be efficient?
5. A local drama company proposes a new neighborhood theater in San Francisco. Before approving the permit, the city planner completes a study of the theater’s impact on the surrounding community.
   a. One finding of the study is that theaters attract traffic, which adversely affects the community. The city planner estimates that the cost to the community from the extra traffic is $5 per ticket. What kind of an externality is this? Why?
   b. Graph the market for theater tickets, labeling the demand curve, the social-value curve, the supply curve, the social-cost curve, the market equilibrium level of output, and the efficient level of output. Also show the per-unit amount of the externality.
   c. Upon further review, the city planner uncovers a second externality. Rehearsals for the plays tend to run until late at night, with actors, stagehands, and other theater members coming and going at various hours. The planner has found that the increased foot traffic improves the safety of the surrounding streets, an estimated benefit to the community of $2 per ticket. What kind of externality is this? Why?
d. On a new graph, illustrate the market for theater tickets in the case of these two externalities. Again, label the demand curve, the social-value curve, the supply curve, the social-cost curve, the market equilibrium level of output, and the per-unit amount of both externalities.
e. Describe a government policy that would result in an efficient outcome.

6. Greater consumption of alcohol leads to more motor vehicle accidents and, thus, imposes costs on people who do not drink and drive.
a. Illustrate the market for alcohol, labeling the demand curve, the social-value curve, the supply curve, the social-cost curve, the market equilibrium level of output, and the efficient level of output.
b. On your graph, shade the area corresponding to the deadweight loss of the market equilibrium. (Hint: The deadweight loss occurs because some units of alcohol are consumed for which the social cost exceeds the social value.) Explain.

7. Many observers believe that the levels of pollution in our society are too high.
a. If society wishes to reduce overall pollution by a certain amount, why is it efficient to have different amounts of reduction at different firms?
b. Command-and-control approaches often rely on uniform reductions among firms. Why are these approaches generally unable to target the firms that should undertake bigger reductions?
c. Economists argue that appropriate corrective taxes or tradable pollution rights will result in efficient pollution reduction. How do these approaches target the firms that should undertake bigger reductions?

8. Ringo loves playing rock-'n'-roll music at high volume. Luciano loves opera and hates rock-'n'-roll. Unfortunately, they are next-door neighbors in an apartment building with paper-thin walls.
a. What is the externality here?
b. What command-and-control policy might the landlord impose? Could such a policy lead to an inefficient outcome?
c. Suppose the landlord lets the tenants do whatever they want. According to the Coase theorem, how might Ringo and Luciano reach an efficient outcome on their own? What might prevent them from reaching an efficient outcome?

9. The Pristine River has two polluting firms on its banks. Acme Industrial and Creative Chemicals each dump 100 tons of glop into the river each year. The cost of reducing glorp emissions per ton equals $10 for Acme and $100 for Creative. The local government wants to reduce overall pollution from 200 tons to 50 tons.
a. If the government knows the cost of reduction for each firm, what reductions will it impose to reach its overall goal? What will be the cost to each firm and the total cost to the firms together?
b. In a more typical situation, the government does not know the cost of pollution reduction for each firm. If the government decides to reach its overall goal by imposing uniform reductions on the firms, calculate the reduction made by each firm, the cost to each firm, and the total cost to the firms together.
c. Compare the total cost of pollution reduction in parts (a) and (b). If the government does not know the cost of reduction for each firm, is there still some way for it to reduce pollution to 50 tons at the total cost you calculated in part (a)? Explain.

10. Figure 4 shows that for any given demand curve for the right to pollute, the government can achieve the same outcome either by setting a price with a corrective tax or by setting a quantity with pollution permits. Suppose there is a sharp improvement in the technology for controlling pollution.
a. Using graphs similar to those in Figure 4, illustrate the effect of this development on the demand for pollution rights.
b. What is the effect on the price and quantity of pollution under each regulatory system? Explain.

11. Suppose that the government decides to issue tradable permits for a certain form of pollution.
a. Does it matter for economic efficiency whether the government distributes or auctions the permits?
b. If the government chooses to distribute the permits, does the allocation of permits among firms matter for efficiency?
b. For each unit of Negext produced, 4 units of pollution are emitted, and each unit of pollution imposes a cost on society of $1. Compute the total cost of pollution when the market for Negext is in equilibrium. What is total surplus from this market after taking into account the cost of pollution?

c. Would banning Negext increase or decrease welfare? Why?

d. Suppose that the government restricts emissions to 100 units of pollution. Graph the Negext market under this constraint. Find the new equilibrium price and quantity and show them on your graph. Compute how this policy affects consumer surplus, producer surplus, and the cost of pollution. Would you recommend this policy? Why?

e. Suppose that instead of restricting pollution, the government imposes a tax on producers equal to $4 for each unit of chemical produced. Calculate the new equilibrium price and quantity, as well as consumer surplus, producer surplus, tax revenue, and the cost of pollution. What is total surplus now? Would you recommend this policy? Why?

f. New research finds the social cost of pollution is really higher than $1. How would that change the optimal policy response? Is there some cost of pollution that would make it sensible to ban Negext? If so, what is it?
Public Goods and Common Resources

An old song lyric maintains that “the best things in life are free.” A moment’s thought reveals a long list of goods that the songwriter could have had in mind. Nature provides some of them, such as rivers, mountains, beaches, lakes, and oceans. The government provides others, such as playgrounds, parks, and parades. In each case, people do not pay a fee when they choose to enjoy the benefit of the good.

Goods without prices provide a special challenge for economic analysis. Most goods in our economy are allocated in markets, where buyers pay for what they receive and sellers are paid for what they provide. For these goods, prices are the signals that guide the decisions of buyers and sellers, and these decisions lead to an efficient allocation of resources. When goods are available free of charge, however, the market forces that normally allocate resources in our economy are absent.

In this chapter, we examine the problems that arise for the allocation of resources when there are goods without market prices. Our analysis will shed light on one of the Ten Principles of Economics in Chapter 1: Governments can sometimes improve market outcomes. When a good does not have a price attached to it, private markets cannot ensure that the good is produced and consumed in the proper amounts. In such cases, government policy can potentially remedy the market failure and raise economic well-being.
The Different Kinds of Goods

How well do markets work in providing the goods that people want? The answer to this question depends on the good being considered. As we discussed in Chapter 7, a market can provide the efficient number of ice-cream cones: The price of ice-cream cones adjusts to balance supply and demand, and this equilibrium maximizes the sum of producer and consumer surplus. Yet as we discussed in Chapter 10, the market cannot be counted on to prevent aluminum manufacturers from polluting the air we breathe: Buyers and sellers in a market typically do not take into account the external effects of their decisions. Thus, markets work well when the good is ice cream, but they work badly when the good is clean air.

In thinking about the various goods in the economy, it is useful to group them according to two characteristics:

- Is the good **excludable**? That is, can people be prevented from using the good?
- Is the good **rival in consumption**? That is, does one person’s use of the good reduce another person’s ability to use it?

Using these two characteristics, Figure 1 divides goods into four categories:

1. **Private goods** are both excludable and rival in consumption. Consider an ice-cream cone, for example. An ice-cream cone is excludable because it is possible to prevent someone from eating an ice-cream cone—you just don’t give it to him. An ice-cream cone is rival in consumption because if one person eats an ice-cream cone, another person cannot eat the same cone. Most goods in the economy are private goods like ice-cream cones: You don’t get one unless you pay, and once you have it, you are the only person who benefits. When we analyzed supply and demand in Chapters 4, 5, and 6 and the efficiency of markets in Chapters 7, 8, and 9, we implicitly assumed that goods were both excludable and rival in consumption.

2. **Public goods** are neither excludable nor rival in consumption. That is, people cannot be prevented from using a public good, and one person’s use of a public good does not reduce another person’s ability to use it. For example,
a tornado siren in a small town is a public good. Once the siren sounds, it is impossible to prevent any single person from hearing it (so it is not excludable). Moreover, when one person gets the benefit of the warning, she does not reduce the benefit to anyone else (so it is not rival in consumption).

3. **Common resources** are rival in consumption but not excludable. For example, fish in the ocean are rival in consumption: When one person catches fish, there are fewer fish for the next person to catch. Yet these fish are not an excludable good because, given the vast size of an ocean, it is difficult to stop fishermen from taking fish out of it.

4. When a good is excludable but not rival in consumption, it is an example of a good produced by a natural monopoly. For instance, consider fire protection in a small town. It is easy to exclude someone from using this good: The fire department can just let his house burn down. Yet fire protection is not rival in consumption: Once a town has paid for the fire department, the additional cost of protecting one more house is small. (In Chapter 15, we give a more complete definition of natural monopolies and study them in some detail.)

Although Figure 1 offers a clean separation of goods into four categories, the boundary between the categories is sometimes fuzzy. Whether goods are excludable or rival in consumption is often a matter of degree. Fish in an ocean may not be excludable because monitoring fishing is so difficult, but a large enough coast guard could make fish at least partly excludable. Similarly, although fish are generally rival in consumption, this would be less true if the population of fishermen were small relative to the population of fish. (Think of North American fishing waters before the arrival of European settlers.) For purposes of our analysis, however, it will be helpful to group goods into these four categories.

In this chapter, we examine goods that are not excludable: public goods and common resources. Because people cannot be prevented from using these goods, they are available to everyone free of charge. The study of public goods and common resources is closely related to the study of externalities. For both of these types of goods, externalities arise because something of value has no price attached to it. If one person were to provide a public good, such as a tornado siren, other people would be better off. They would receive a benefit without paying for it—a positive externality. Similarly, when one person uses a common resource such as the fish in the ocean, other people are worse off because there are fewer fish to catch. They suffer a loss but are not compensated for it—a negative externality. Because of these external effects, private decisions about consumption and production can lead to an inefficient allocation of resources, and government intervention can potentially raise economic well-being.

**Quick Quiz** Define public goods and common resources and give an example of each.

---

**PUBLIC GOODS**

To understand how public goods differ from other goods and the problems they present for society, let’s consider an example: a fireworks display. This good is not excludable because it is impossible to prevent someone from seeing fireworks, and it is not rival in consumption because one person’s enjoyment of fireworks does not reduce anyone else’s enjoyment of them.
The Free-Rider Problem

The citizens of Smalltown, U.S.A., like seeing fireworks on the Fourth of July. Each of the town’s 500 residents places a $10 value on the experience for a total benefit of $5,000. The cost of putting on a fireworks display is $1,000. Because the $5,000 benefit exceeds the $1,000 cost, it is efficient for Smalltown to have a fireworks display on the Fourth of July.

Would the private market produce the efficient outcome? Probably not. Imagine that Ellen, a Smalltown entrepreneur, decided to put on a fireworks display. Ellen would surely have trouble selling tickets to the event because her potential customers would quickly figure out that they could see the fireworks even without a ticket. Because fireworks are not excludable, people have an incentive to be free riders. A free rider is a person who receives the benefit of a good but does not pay for it. Because people would have an incentive to be free riders rather than ticket buyers, the market would fail to provide the efficient outcome.

One way to view this market failure is that it arises because of an externality. If Ellen puts on the fireworks display, she confers an external benefit on those who see the display without paying for it. When deciding whether to put on the display, however, Ellen does not take the external benefits into account. Even though the fireworks display is socially desirable, it is not profitable. As a result, Ellen makes the privately rational but socially inefficient decision not to put on the display.

Although the private market fails to supply the fireworks display demanded by Smalltown residents, the solution to Smalltown’s problem is obvious: The local government can sponsor a Fourth of July celebration. The town council can raise everyone’s taxes by $2 and use the revenue to hire Ellen to produce the fireworks. Everyone in Smalltown is better off by $8—the $10 in value from the fireworks minus the $2 tax bill. Ellen can help Smalltown reach the efficient outcome as a public employee even though she could not do so as a private entrepreneur.

The story of Smalltown is simplified but realistic. In fact, many local governments in the United States pay for fireworks on the Fourth of July. Moreover, the story shows a general lesson about public goods: Because public goods are not excludable, the free-rider problem prevents the private market from supplying them. The government, however, can potentially remedy the problem. If the government decides that the total benefits of a public good exceed its costs, it can provide the public good, pay for it with tax revenue, and make everyone better off.

Some Important Public Goods

There are many examples of public goods. Here we consider three of the most important.

National Defense The defense of a country from foreign aggressors is a classic example of a public good. Once the country is defended, it is impossible to prevent any single person from enjoying the benefit of this defense. Moreover, when one person enjoys the benefit of national defense, he does not reduce the benefit to anyone else. Thus, national defense is neither excludable nor rival in consumption.

National defense is also one of the most expensive public goods. In 2007, the U.S. federal government spent a total of $553 billion on national defense, more than $1,800 per person. People disagree about whether this amount is too small or
too large, but almost no one doubts that some government spending for national defense is necessary. Even economists who advocate small government agree that the national defense is a public good the government should provide.

**Basic Research** Knowledge is created through research. In evaluating the appropriate public policy toward knowledge creation, it is important to distinguish general knowledge from specific technological knowledge. Specific technological knowledge, such as the invention of a longer-lasting battery, a smaller microchip, or a better digital music player, can be patented. The patent gives the inventor the exclusive right to the knowledge he or she has created for a period of time. Anyone else who wants to use the patented information must pay the inventor for the right to do so. In other words, the patent makes the knowledge created by the inventor excludable.

By contrast, general knowledge is a public good. For example, a mathematician cannot patent a theorem. Once a theorem is proved, the knowledge is not excludable: The theorem enters society’s general pool of knowledge that anyone can use without charge. The theorem is also not rival in consumption: One person’s use of the theorem does not prevent any other person from using the theorem.

Profit-seeking firms spend a lot on research trying to develop new products that they can patent and sell, but they do not spend much on basic research. Their incentive, instead, is to free ride on the general knowledge created by others. As a result, in the absence of any public policy, society would devote too few resources to creating new knowledge.

The government tries to provide the public good of general knowledge in various ways. Government agencies, such as the National Institutes of Health and the National Science Foundation, subsidize basic research in medicine, mathematics, physics, chemistry, biology, and even economics. Some people justify government funding of the space program on the grounds that it adds to society’s pool of knowledge (although many scientists are skeptical of the scientific value of manned space travel). Determining the appropriate level of government support for these endeavors is difficult because the benefits are hard to measure. Moreover, the members of Congress who appropriate funds for research usually have little expertise in science and, therefore, are not in the best position to judge what lines of research will produce the largest benefits. So, while basic research is surely a public good, we should not be surprised if the public sector fails to pay for the right amount and the right kinds.

**Fighting Poverty** Many government programs are aimed at helping the poor. The welfare system (officially called Temporary Assistance for Needy Families) provides a small income for some poor families. Similarly, the Food Stamp program subsidizes the purchase of food for those with low incomes, and various government housing programs make shelter more affordable. These antipoverty programs are financed by taxes paid by families that are financially more successful.

Economists disagree among themselves about what role the government should play in fighting poverty. Although we discuss this debate more fully in Chapter 20, here we note one important argument: Advocates of antipoverty programs claim that fighting poverty is a public good. Even if everyone prefers living in a society without poverty, fighting poverty is not a “good” that private actions will adequately provide.
To see why, suppose someone tried to organize a group of wealthy individuals to try to eliminate poverty. They would be providing a public good. This good would not be rival in consumption: One person’s enjoyment of living in a society without poverty would not reduce anyone else’s enjoyment of it. The good would not be excludable: Once poverty is eliminated, no one can be prevented from taking pleasure in this fact. As a result, there would be a tendency for people to free ride on the generosity of others, enjoying the benefits of poverty elimination without contributing to the cause.

Because of the free-rider problem, eliminating poverty through private charity will probably not work. Yet government action can solve this problem. Taxing the wealthy to raise the living standards of the poor can potentially make everyone better off. The poor are better off because they now enjoy a higher standard of living, and those paying the taxes are better off because they enjoy living in a society with less poverty.

ARE LIGHTHOUSES PUBLIC GOODS?

Some goods can switch between being public goods and being private goods depending on the circumstances. For example, a fireworks display is a public good if performed in a town with many residents. Yet if performed at a private amusement park, such as Walt Disney World, a fireworks display is more like a private good because visitors to the park pay for admission.

Another example is a lighthouse. Economists have long used lighthouses as an example of a public good. Lighthouses mark specific locations so that passing ships can avoid treacherous waters. The benefit that the lighthouse provides to the ship captain is neither excludable nor rival in consumption, so each captain has an incentive to free ride by using the lighthouse to navigate without paying for the service. Because of this free-rider problem, private markets usually fail to provide the lighthouses that ship captains need. As a result, most lighthouses today are operated by the government.

In some cases, however, lighthouses have been closer to private goods. On the coast of England in the 19th century, for example, some lighthouses were privately owned and operated. Instead of trying to charge ship captains for the service, however, the owner of the lighthouse charged the owner of the nearby port. If the port owner did not pay, the lighthouse owner turned off the light, and ships avoided that port.

In deciding whether something is a public good, one must determine who the beneficiaries are and whether these beneficiaries can be excluded from using the good. A free-rider problem arises when the number of beneficiaries is large and exclusion of any one of them is impossible. If a lighthouse benefits many ship captains, it is a public good. Yet if it primarily benefits a single port owner, it is more like a private good.

THE DIFFICULT JOB OF COST–BENEFIT ANALYSIS

So far we have seen that the government provides public goods because the private market on its own will not produce an efficient quantity. Yet deciding that the government must play a role is only the first step. The government must then determine what kinds of public goods to provide and in what quantities.
Suppose that the government is considering a public project, such as building a new highway. To judge whether to build the highway, it must compare the total benefits of all those who would use it to the costs of building and maintaining it. To make this decision, the government might hire a team of economists and engineers to conduct a study, called a cost–benefit analysis, the goal of which is to estimate the total costs and benefits of the project to society as a whole.

Cost–benefit analysts have a tough job. Because the highway will be available to everyone free of charge, there is no price with which to judge the value of the highway. Simply asking people how much they would value the highway is not reliable: Quantifying benefits is difficult using the results from a questionnaire, and respondents have little incentive to tell the truth. Those who would use the highway have an incentive to exaggerate the benefit they receive to get the highway built. Those who would be harmed by the highway have an incentive to exaggerate the costs to them to prevent the highway from being built.

The efficient provision of public goods is, therefore, intrinsically more difficult than the efficient provision of private goods. When buyers of a private good enter a market, they reveal the value they place on it through the prices they are willing to pay. At the same time, sellers reveal their costs with the prices they are willing to accept. The equilibrium is an efficient allocation of resources because it reflects all this information. By contrast, cost–benefit analysts do not have any price signals to observe when evaluating whether the government should provide a public good and how much to provide. Their findings on the costs and benefits of public projects are rough approximations at best.

**HOW MUCH IS A LIFE WORTH?**

Imagine that you have been elected to serve as a member of your local town council. The town engineer comes to you with a proposal: The town can spend $10,000 to build and operate a traffic light at a town intersection that now has only a stop sign. The benefit of the traffic light is increased safety. The engineer estimates, based on data from similar intersections, that the traffic light would reduce the risk of a fatal traffic accident over the lifetime of the traffic light from 1.6 to 1.1 percent. Should you spend the money for the new light?

To answer this question, you turn to cost–benefit analysis. But you quickly run into an obstacle: The costs and benefits must be measured in the same units if you are to compare them meaningfully. The cost is measured in dollars, but the benefit—the possibility of saving a person’s life—is not directly monetary. To make your decision, you have to put a dollar value on a human life.

At first, you may be tempted to conclude that a human life is priceless. After all, there is probably no amount of money that you could be paid to voluntarily give up your life or that of a loved one. This suggests that a human life has an infinite dollar value.

For the purposes of cost–benefit analysis, however, this answer leads to nonsensical results. If we truly placed an infinite value on human life, we should place traffic lights on every street corner, and we should all drive large cars loaded with all the latest safety features. Yet traffic lights are not at every corner, and people sometimes choose to pay less for smaller cars without safety options such as side-impact air bags or antilock brakes. In both our public and private decisions, we are at times willing to risk our lives to save some money.
Once we have accepted the idea that a person’s life has an implicit dollar value, how can we determine what that value is? One approach, sometimes used by courts to award damages in wrongful-death suits, is to look at the total amount of money a person would have earned if he or she had lived. Economists are often critical of this approach because it ignores other opportunity costs of losing one’s life. It thus has the bizarre implication that the life of a retired or disabled person has no value.

A better way to value human life is to look at the risks that people are voluntarily willing to take and how much they must be paid for taking them. Mortality risk varies across jobs, for example. Construction workers in high-rise buildings face greater risk of death on the job than office workers do. By comparing wages in risky and less risky occupations, controlling for education, experience, and other determinants of wages, economists can get some sense about what value people put on their own lives. Studies using this approach conclude that the value of a human life is about $10 million.

We can now return to our original example and respond to the town engineer. The traffic light reduces the risk of fatality by 0.5 percentage points. Thus, the expected benefit from installing the traffic light is $0.005 \times $10 million, or $50,000. This estimate of the benefit well exceeds the cost of $10,000, so you should approve the project.

**Quick Quiz**

What is the *free-rider problem*? Why does the free-rider problem induce the government to provide public goods? How should the government decide whether to provide a public good?

---

**Common Resources**

Common resources, like public goods, are not excludable: They are available free of charge to anyone who wants to use them. Common resources are, however, rival in consumption: One person’s use of the common resource reduces other people’s ability to use it. Thus, common resources give rise to a new problem. Once the good is provided, policymakers need to be concerned about how much it is used. This problem is best understood from the classic parable called the *Tragedy of the Commons*.

**The Tragedy of the Commons**

Consider life in a small medieval town. Of the many economic activities that take place in the town, one of the most important is raising sheep. Many of the town’s families own flocks of sheep and support themselves by selling the sheep’s wool, which is used to make clothing.

As our story begins, the sheep spend much of their time grazing on the land surrounding the town, called the Town Common. No family owns the land. Instead, the town residents own the land collectively, and all the residents are allowed to graze their sheep on it. Collective ownership works well because land is plentiful. As long as everyone can get all the good grazing land they want, the Town Common is not rival in consumption, and allowing residents’ sheep to graze for free causes no problems. Everyone in town is happy.
As the years pass, the population of the town grows, and so does the number of sheep grazing on the Town Common. With a growing number of sheep and a fixed amount of land, the land starts to lose its ability to replenish itself. Eventually, the land is grazed so heavily that it becomes barren. With no grass left on the Town Common, raising sheep is impossible, and the town’s once prosperous wool industry disappears. Many families lose their source of livelihood.

What causes the tragedy? Why do the shepherds allow the sheep population to grow so large that it destroys the Town Common? The reason is that social and private incentives differ. Avoiding the destruction of the grazing land depends on the collective action of the shepherds. If the shepherds acted together, they could reduce the sheep population to a size that the Town Common can support. Yet no single family has an incentive to reduce the size of its own flock because each flock represents only a small part of the problem.

In essence, the Tragedy of the Commons arises because of an externality. When one family’s flock grazes on the common land, it reduces the quality of the land available for other families. Because people neglect this negative externality when deciding how many sheep to own, the result is an excessive number of sheep.

If the tragedy had been foreseen, the town could have solved the problem in various ways. It could have regulated the number of sheep in each family’s flock, internalized the externality by taxing sheep, or auctioned off a limited number of sheep-grazing permits. That is, the medieval town could have dealt with the problem of overgrazing in the way that modern society deals with the problem of pollution.

In the case of land, however, there is a simpler solution. The town can divide the land among town families. Each family can enclose its parcel of land with a fence and then protect it from excessive grazing. In this way, the land becomes a private good rather than a common resource. This outcome in fact occurred during the enclosure movement in England in the 17th century.

The Tragedy of the Commons is a story with a general lesson: When one person uses a common resource, he or she diminishes other people’s enjoyment of it. Because of this negative externality, common resources tend to be used excessively. The government can solve the problem by using regulation or taxes to reduce consumption of the common resource. Alternatively, the government can sometimes turn the common resource into a private good.

This lesson has been known for thousands of years. The ancient Greek philosopher Aristotle pointed out the problem with common resources: “What is common to many is taken least care of, for all men have greater regard for what is their own than for what they possess in common with others.”

**Some Important Common Resources**

There are many examples of common resources. In almost all cases, the same problem arises as in the Tragedy of the Commons: Private decision makers use the common resource too much. Governments often regulate behavior or impose fees to mitigate the problem of overuse.

**Clean Air and Water** As we discussed in Chapter 10, markets do not adequately protect the environment. Pollution is a negative externality that can be remedied with regulations or with corrective taxes on polluting activities. One can view this market failure as an example of a common-resource problem. Clean air and clean
water are common resources like open grazing land, and excessive pollution is like excessive grazing. Environmental degradation is a modern Tragedy of the Commons.

**Congested Roads**  
Roads can be either public goods or common resources. If a road is not congested, then one person’s use does not affect anyone else. In this case, use is not rival in consumption, and the road is a public good. Yet if a road is congested, then use of that road yields a negative externality. When one person drives on the road, it becomes more crowded, and other people must drive more slowly. In this case, the road is a common resource.
One way for the government to address the problem of road congestion is to charge drivers a toll. A toll is, in essence, a corrective tax on the externality of congestion. Sometimes, as in the case of local roads, tolls are not a practical solution because the cost of collecting them is too high. But the city of London has found increasing tolls to be a very effective way to reduce congestion, and as the accompanying In The News box discusses, a similar plan is being considered for New York City.

Sometimes congestion is a problem only at certain times of day. If a bridge is heavily traveled only during rush hour, for instance, the congestion externality is largest during this time. The efficient way to deal with these externalities is to
charge higher tolls during rush hour. This toll would provide an incentive for drivers to alter their schedules, reducing traffic when congestion is greatest.

Another policy that responds to the problem of road congestion, discussed in a case study in the previous chapter, is the tax on gasoline. Gasoline is a complementary good to driving: An increase in the price of gasoline tends to reduce the quantity of driving demanded. Therefore, a gasoline tax reduces road congestion. A gasoline tax, however, is an imperfect solution, because it affects other decisions besides the amount of driving on congested roads. For example, the gasoline tax discourages driving on uncongested roads, even though there is no congestion externality for these roads.

Fish, Whales, and Other Wildlife  Many species of animals are common resources. Fish and whales, for instance, have commercial value, and anyone can go to the ocean and catch whatever is available. Each person has little incentive to maintain the species for the next year. Just as excessive grazing can destroy the Town Common, excessive fishing and whaling can destroy commercially valuable marine populations.

The ocean remains one of the least regulated common resources. Two problems prevent an easy solution. First, many countries have access to the oceans, so any solution would require international cooperation among countries that hold different values. Second, because the oceans are so vast, enforcing any agreement is difficult. As a result, fishing rights have been a frequent source of international tension among normally friendly countries.

Within the United States, various laws aim to protect fish and other wildlife. For example, the government charges for fishing and hunting licenses, and it restricts the lengths of the fishing and hunting seasons. Fishermen are often required to throw back small fish, and hunters can kill only a limited number of animals. All these laws reduce the use of a common resource and help maintain animal populations.

**WHY THE COW IS NOT EXTINCT**

Throughout history, many species of animals have been threatened with extinction. When Europeans first arrived in North America, more than 60 million buffalo roamed the continent. Yet hunting the buffalo was so popular during the 19th century that by 1900 the animal’s population had fallen to about 400 before the government stepped in to protect the species. In some African countries today, the elephant faces a similar challenge, as poachers kill the animals for the ivory in their tusks.

Yet not all animals with commercial value face this threat. The cow, for example, is a valuable source of food, but no one worries that the cow will soon be extinct. Indeed, the great demand for beef seems to ensure that the species will continue to thrive.

Why is the commercial value of ivory a threat to the elephant, while the commercial value of beef is a guardian of the cow? The reason is that elephants are a common resource, whereas cows are a private good. Elephants roam freely with-
out any owners. Each poacher has a strong incentive to kill as many elephants as he can find. Because poachers are numerous, each poacher has only a slight incentive to preserve the elephant population. By contrast, cattle live on ranches that are privately owned. Each rancher makes great effort to maintain the cattle population on his ranch because he reaps the benefit of these efforts.

Governments have tried to solve the elephant’s problem in two ways. Some countries, such as Kenya, Tanzania, and Uganda, have made it illegal to kill elephants and sell their ivory. Yet these laws have been hard to enforce, and elephant populations have continued to dwindle. By contrast, other countries, such as Botswana, Malawi, Namibia, and Zimbabwe, have made elephants a private good by allowing people to kill elephants, but only those on their own property. Landowners now have an incentive to preserve the species on their own land, and as a result, elephant populations have started to rise. With private ownership and the profit motive now on its side, the African elephant might someday be as safe from extinction as the cow.

**QUICK QUIZ** Why do governments try to limit the use of common resources?

**CONCLUSION: THE IMPORTANCE OF PROPERTY RIGHTS**

In this and the previous chapter, we have seen there are some “goods” that the market does not provide adequately. Markets do not ensure that the air we breathe is clean or that our country is defended from foreign aggressors. Instead, societies rely on the government to protect the environment and to provide for the national defense.

Although the problems we considered in these chapters arise in many different markets, they share a common theme. In all cases, the market fails to allocate resources efficiently because *property rights* are not well established. That is, some item of value does not have an owner with the legal authority to control it. For example, although no one doubts that the “good” of clean air or national defense is valuable, no one has the right to attach a price to it and profit from its use. A factory pollutes too much because no one charges the factory for the pollution it emits. The market does not provide for national defense because no one can charge those who are defended for the benefit they receive.

When the absence of property rights causes a market failure, the government can potentially solve the problem. Sometimes, as in the sale of pollution permits, the solution is for the government to help define property rights and thereby unleash market forces. Other times, as in restricted hunting seasons, the solution is for the government to regulate private behavior. Still other times, as in the provision of national defense, the solution is for the government to use tax revenue to supply a good that the market fails to supply. In all cases, if the policy is well planned and well run, it can make the allocation of resources more efficient and thus raise economic well-being.
SUMMARY

- Goods differ in whether they are excludable and whether they are rival in consumption. A good is excludable if it is possible to prevent someone from using it. A good is rival in consumption if one person’s use of the good reduces other people’s ability to use the same unit of the good. Markets work best for private goods, which are both excludable and rival in consumption. Markets do not work as well for other types of goods.
- Public goods are neither rival in consumption nor excludable. Examples of public goods include fireworks displays, national defense, and the creation of fundamental knowledge. Because people are not charged for their use of the public good, they have an incentive to free ride when the good is provided privately. Therefore, governments provide public goods, making their decision about the quantity of each good based on cost–benefit analysis.
- Common resources are rival in consumption but not excludable. Examples include common grazing land, clean air, and congested roads. Because people are not charged for their use of common resources, they tend to use them excessively. Therefore, governments use various methods to limit the use of common resources.

KEY CONCEPTS


QUESTIONS FOR REVIEW

1. Explain what is meant by a good being “excludable.” Explain what is meant by a good being “rival in consumption.” Is a slice of pizza excludable? Is it rival in consumption?
2. Define and give an example of a public good. Can the private market provide this good on its own? Explain.
3. What is cost–benefit analysis of public goods? Why is it important? Why is it hard?
4. Define and give an example of a common resource. Without government intervention, will people use this good too much or too little? Why?

PROBLEMS AND APPLICATIONS

1. Think about the goods and services provided by your local government.
   a. Using the classification in Figure 1, explain which category each of the following goods falls into:
      • police protection
      • snow plowing
   b. Why do you think the government provides items that are not public goods?
2. Both public goods and common resources involve externalities.
a. Are the externalities associated with public goods generally positive or negative? Use examples in your answer. Is the free-market quantity of public goods generally greater or less than the efficient quantity?
b. Are the externalities associated with common resources generally positive or negative? Use examples in your answer. Is the free-market use of common resources generally greater or less than the efficient use?

3. Charlie loves watching *Teletubbies* on his local public TV station, but he never sends any money to support the station during its fundraising drives.
   a. What name do economists have for Charlie?
   b. How can the government solve the problem caused by people like Charlie?
   c. Can you think of ways the private market can solve this problem? How does the existence of cable TV alter the situation?

4. Four roommates are planning to spend the weekend in their dorm room watching old movies, and they are debating how many to watch. Here is their willingness to pay for each film:

<table>
<thead>
<tr>
<th></th>
<th>Orson</th>
<th>Alfred</th>
<th>Woody</th>
<th>Ingmar</th>
</tr>
</thead>
<tbody>
<tr>
<td>First film</td>
<td>$7</td>
<td>$5</td>
<td>$3</td>
<td>$2</td>
</tr>
<tr>
<td>Second film</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Third film</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Fourth film</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fifth film</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

   a. Within the dorm room, is the showing of a movie a public good? Why or why not?
   b. If it costs $8 to rent a movie, how many movies should the roommates rent to maximize total surplus?
   c. If they choose the optimal number from part (b) and then split the cost of renting the movies equally, how much surplus does each person obtain from watching the movies?
   d. Is there any way to split the cost to ensure that everyone benefits? What practical problems does this solution raise?
   e. Suppose they agree in advance to choose the efficient number and to split the cost of the movies equally. When Orson is asked his willingness to pay, will he have an incentive to tell the truth? If so, why? If not, what will he be tempted to say?

f. What does this example teach you about the optimal provision of public goods?

5. Some economists argue that private firms will not undertake the efficient amount of basic scientific research.
   a. Explain why this might be so. In your answer, classify basic research in one of the categories shown in Figure 1.
   b. What sort of policy has the United States adopted in response to this problem?
   c. It is often argued that this policy increases the technological capability of American producers relative to that of foreign firms. Is this argument consistent with your classification of basic research in part (a)? (Hint: Can excludability apply to some potential beneficiaries of a public good and not others?)

6. There is often litter along highways but rarely in people’s yards. Provide an economic explanation for this fact.

7. The village of Ectenia has ten residents. Villagers can earn income by either weaving baskets or fishing. Because the lake has a limited number of fish, the more villagers fish, the less each catches. In particular, if \( n \) households fish in the lake, then each fishing household makes an amount:

   \[ I_f = 12 - 2n \]

   where \( I_f \) is daily income measured in dollars. The income that a household makes by weaving baskets is $2 a day.

   a. Assume that each household makes the decision of whether to weave baskets or fish in the lake independently. How many households do you expect to see fishing each day? How many households do you expect to see weaving baskets? (Hint: Think about opportunity cost.) Calculate the total income of the village in this equilibrium.
   b. Show that, when 3 households fish in the lake, the total income of the village is larger than the one you found in part (a). What prevented the villagers from reaching this higher-income allocation of resources when they acted independently?
   c. If the villagers together decided to achieve the allocation in part (b), what kinds of rules would they need to institute? If they wanted
10. The federal government tests the safety of car models and provides the test results free of charge to the public. Do you think this information qualifies as a public good? Why or why not?

11. High-income people are willing to pay more than lower-income people to avoid the risk of death. For example, they are more likely to pay for safety features on cars. Do you think cost–benefit analysts should take this fact into account when evaluating public projects? Consider, for instance, a rich town and a poor town, both of which are considering the installation of a traffic light. Should the rich town use a higher dollar value for a human life in making this decision? Why or why not?
PART V

Firm Behavior and the Organization of Industry
This page intentionally left blank
The Costs of Production

The economy is made up of thousands of firms that produce the goods and services you enjoy every day: General Motors produces automobiles, General Electric produces lightbulbs, and General Mills produces breakfast cereals. Some firms, such as these three, are large; they employ thousands of workers and have thousands of stockholders who share in the firms’ profits. Other firms, such as the local barbershop or candy store, are small; they employ only a few workers and are owned by a single person or family.

In previous chapters, we used the supply curve to summarize firms’ production decisions. According to the law of supply, firms are willing to produce and sell a greater quantity of a good when the price of the good is higher, and this response leads to a supply curve that slopes upward. For analyzing many questions, the law of supply is all you need to know about firm behavior.

In this chapter and the ones that follow, we examine firm behavior in more detail. This topic will give you a better understanding of the decisions behind the supply curve. In addition, it will introduce you to a part of economics called industrial organization—the study of how firms’ decisions about prices and quantities depend on the market conditions they face. The town in which you live, for instance, may have several pizzerias but only one cable television company. This raises a key question: How does the number of firms affect the prices in a market and the efficiency of the market outcome? The field of industrial organization addresses exactly this question.
Before turning to these issues, we need to discuss the costs of production. All firms, from Delta Air Lines to your local deli, incur costs as they make the goods and services that they sell. As we will see in the coming chapters, a firm’s costs are a key determinant of its production and pricing decisions. In this chapter, we define some of the variables that economists use to measure a firm’s costs, and we consider the relationships among these variables.

A word of warning: This topic is dry and technical. To be honest, one might even call it boring. But this material provides a crucial foundation for the fascinating topics that follow.

WHAT ARE COSTS?

We begin our discussion of costs at Caroline’s Cookie Factory. Caroline, the owner of the firm, buys flour, sugar, chocolate chips, and other cookie ingredients. She also buys the mixers and ovens and hires workers to run this equipment. She then sells the cookies to consumers. By examining some of the issues that Caroline faces in her business, we can learn some lessons about costs that apply to all firms in an economy.

TOTAL REVENUE, TOTAL COST, AND PROFIT

We begin with the firm’s objective. To understand the decisions a firm makes, we must understand what it is trying to do. It is conceivable that Caroline started her firm because of an altruistic desire to provide the world with cookies or, perhaps, out of love for the cookie business. More likely, Caroline started her business to make money. Economists normally assume that the goal of a firm is to maximize profit, and they find that this assumption works well in most cases.

What is a firm’s profit? The amount that the firm receives for the sale of its output (cookies) is called its total revenue. The amount that the firm pays to buy inputs (flour, sugar, workers, ovens, and so forth) is called its total cost. Caroline gets to keep any revenue that is not needed to cover costs. Profit is a firm’s total revenue minus its total cost:

\[ \text{Profit} = \text{Total revenue} - \text{Total cost}. \]

Caroline’s objective is to make her firm’s profit as large as possible.

To see how a firm goes about maximizing profit, we must consider fully how to measure its total revenue and its total cost. Total revenue is the easy part: It equals the quantity of output the firm produces times the price at which it sells its output. If Caroline produces 10,000 cookies and sells them at $2 a cookie, her total revenue is $20,000. By contrast, the measurement of a firm’s total cost is more subtle.

COSTS AS OPPORTUNITY COSTS

When measuring costs at Caroline’s Cookie Factory or any other firm, it is important to keep in mind one of the Ten Principles of Economics from Chapter 1: The cost of something is what you give up to get it. Recall that the opportunity cost of an item refers to all those things that must be forgone to acquire that item. When economists speak of a firm’s cost of production, they include all the opportunity costs of making its output of goods and services.
While some of a firm’s opportunity costs of production are obvious, others are less so. When Caroline pays $1,000 for flour, that $1,000 is an opportunity cost because Caroline can no longer use that $1,000 to buy something else. Similarly, when Caroline hires workers to make the cookies, the wages she pays are part of the firm’s costs. Because these opportunity costs require the firm to pay out some money, they are called **explicit costs**. By contrast, some of a firm’s opportunity costs, called **implicit costs**, do not require a cash outlay. Imagine that Caroline is skilled with computers and could earn $100 per hour working as a programmer. For every hour that Caroline works at her cookie factory, she gives up $100 in income, and this forgone income is also part of her costs. The total cost of Caroline’s business is the sum of the explicit costs and the implicit costs.

The distinction between explicit and implicit costs highlights an important difference between how economists and accountants analyze a business. Economists are interested in studying how firms make production and pricing decisions. Because these decisions are based on both explicit and implicit costs, economists include both when measuring a firm’s costs. By contrast, accountants have the job of keeping track of the money that flows into and out of firms. As a result, they measure the explicit costs but usually ignore the implicit costs.

The difference between economists and accountants is easy to see in the case of Caroline’s Cookie Factory. When Caroline gives up the opportunity to earn money as a computer programmer, her accountant will not count this as a cost of her cookie business. Because no money flows out of the business to pay for this cost, it never shows up on the accountant’s financial statements. An economist, however, will count the forgone income as a cost because it will affect the decisions that Caroline makes in her cookie business. For example, if Caroline’s wage as a computer programmer rises from $100 to $500 per hour, she might decide that running her cookie business is too costly and choose to shut down the factory to become a full-time computer programmer.

**The Cost of Capital as an Opportunity Cost**

An important implicit cost of almost every business is the opportunity cost of the financial capital that has been invested in the business. Suppose, for instance, that Caroline used $300,000 of her savings to buy her cookie factory from its previous owner. If Caroline had instead left this money deposited in a savings account that pays an interest rate of 5 percent, she would have earned $15,000 per year. To own her cookie factory, therefore, Caroline has given up $15,000 a year in interest income. This forgone $15,000 is one of the implicit opportunity costs of Caroline’s business.

As we have already noted, economists and accountants treat costs differently, and this is especially true in their treatment of the cost of capital. An economist views the $15,000 in interest income that Caroline gives up every year as a cost of her business, even though it is an implicit cost. Caroline’s accountant, however, will not show this $15,000 as a cost because no money flows out of the business to pay for it.

To further explore the difference between economists and accountants, let’s change the example slightly. Suppose now that Caroline did not have the entire $300,000 to buy the factory but, instead, used $100,000 of her own savings and borrowed $200,000 from a bank at an interest rate of 5 percent. Caroline’s accountant, who only measures explicit costs, will now count the $10,000 interest paid on the bank loan every year as a cost because this amount of money now flows out of
the firm. By contrast, according to an economist, the opportunity cost of owning the business is still $15,000. The opportunity cost equals the interest on the bank loan (an explicit cost of $10,000) plus the forgone interest on savings (an implicit cost of $5,000).

**Economic Profit versus Accounting Profit**

Now let’s return to the firm’s objective: profit. Because economists and accountants measure costs differently, they also measure profit differently. An economist measures a firm’s economic profit as the firm’s total revenue minus all the opportunity costs (explicit and implicit) of producing the goods and services sold. An accountant measures the firm’s accounting profit as the firm’s total revenue minus only the firm’s explicit costs.

Figure 1 summarizes this difference. Notice that because the accountant ignores the implicit costs, accounting profit is usually larger than economic profit. For a business to be profitable from an economist’s standpoint, total revenue must cover all the opportunity costs, both explicit and implicit.

Economic profit is an important concept because it is what motivates the firms that supply goods and services. As we will see, a firm making positive economic profit will stay in business. It is covering all its opportunity costs and has some revenue left to reward the firm owners. When a firm is making economic losses (that is, when economic profits are negative), the business owners are failing to earn enough revenue to cover all the costs of production. Unless conditions change, the firm owners will eventually close down the business and exit the industry. To understand business decisions, we need to keep an eye on economic profit.

**Quick Quiz** Farmer McDonald gives banjo lessons for $20 an hour. One day, he spends 10 hours planting $100 worth of seeds on his farm. What opportunity cost has he incurred? What cost would his accountant measure? If these seeds yield $200 worth of crops, does McDonald earn an accounting profit? Does he earn an economic profit?

---

**Figure 1**

**Economists versus Accountants**

Economists include all opportunity costs when analyzing a firm, whereas accountants measure only explicit costs. Therefore, economic profit is smaller than accounting profit.
PRODUCTION AND COSTS

Firms incur costs when they buy inputs to produce the goods and services that they plan to sell. In this section, we examine the link between a firm’s production process and its total cost. Once again, we consider Caroline’s Cookie Factory.

In the analysis that follows, we make an important simplifying assumption: We assume that the size of Caroline’s factory is fixed and that Caroline can vary the quantity of cookies produced only by changing the number of workers she employs. This assumption is realistic in the short run but not in the long run. That is, Caroline cannot build a larger factory overnight, but she can do so over the next year or two. This analysis, therefore, describes the production decisions that Caroline faces in the short run. We examine the relationship between costs and time horizon more fully later in the chapter.

THE PRODUCTION FUNCTION

Table 1 shows how the quantity of cookies produced per hour at Caroline’s factory depends on the number of workers. As you can see in the first two columns, if there are no workers in the factory, Caroline produces no cookies. When there is 1 worker, she produces 50 cookies. When there are 2 workers, she produces 90 cookies and so on. Panel (a) of Figure 2 presents a graph of these two columns of numbers. The number of workers is on the horizontal axis, and the number of cookies produced is on the vertical axis. This relationship between the quantity of inputs (workers) and quantity of output (cookies) is called the production function.

<table>
<thead>
<tr>
<th>Number of Workers</th>
<th>Output (quantity of cookies produced per hour)</th>
<th>Marginal Product of Labor</th>
<th>Cost of Factory</th>
<th>Cost of Workers</th>
<th>Total Cost of Inputs (cost of factory + cost of workers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>50</td>
<td>$30</td>
<td>$0</td>
<td>$30</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>40</td>
<td>30</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>90</td>
<td>30</td>
<td>30</td>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>120</td>
<td>20</td>
<td>30</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>140</td>
<td>10</td>
<td>30</td>
<td>40</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>150</td>
<td>5</td>
<td>30</td>
<td>50</td>
<td>80</td>
</tr>
<tr>
<td>6</td>
<td>155</td>
<td></td>
<td>30</td>
<td>60</td>
<td>90</td>
</tr>
</tbody>
</table>
One of the Ten Principles of Economics introduced in Chapter 1 is that rational people think at the margin. As we will see in future chapters, this idea is the key to understanding the decisions a firm makes about how many workers to hire and how much output to produce. To take a step toward understanding these decisions, the third column in the table gives the marginal product of a worker. The marginal product of any input in the production process is the increase in the quantity of output obtained from one additional unit of that input. When the number of workers goes from 1 to 2, cookie production increases from 50 to 90, so the marginal product of the second worker is 40 cookies. And when the number of workers goes from 2 to 3, cookie production increases from 90 to 120, so the marginal product of the third worker is 30 cookies. In the table, the marginal product is shown halfway between two rows because it represents the change in output as the number of workers increases from one level to another.

The production function in panel (a) shows the relationship between the number of workers hired and the quantity of output produced. Here the number of workers hired (on the horizontal axis) is from the first column in Table 1, and the quantity of output produced (on the vertical axis) is from the second column. The production function gets flatter as the number of workers increases, which reflects diminishing marginal product. The total-cost curve in panel (b) shows the relationship between the quantity of output produced and total cost of production. Here the quantity of output produced (on the horizontal axis) is from the second column in Table 1, and the total cost (on the vertical axis) is from the sixth column. The total-cost curve gets steeper as the quantity of output increases because of diminishing marginal product.

**marginal product**
the increase in output that arises from an additional unit of input
Notice that as the number of workers increases, the marginal product declines. The second worker has a marginal product of 40 cookies, the third worker has a marginal product of 30 cookies, and the fourth worker has a marginal product of 20 cookies. This property is called diminishing marginal product. At first, when only a few workers are hired, they have easy access to Caroline’s kitchen equipment. As the number of workers increases, additional workers have to share equipment and work in more crowded conditions. Eventually, the kitchen is so crowded that the workers start getting in each other’s way. Hence, as more and more workers are hired, each additional worker contributes fewer additional cookies to total production.

Diminishing marginal product is also apparent in Figure 2. The production function’s slope (“rise over run”) tells us the change in Caroline’s output of cookies (“rise”) for each additional input of labor (“run”). That is, the slope of the production function measures the marginal product of a worker. As the number of workers increases, the marginal product declines, and the production function becomes flatter.

**From the Production Function to the Total-Cost Curve**

The last three columns of Table 1 show Caroline’s cost of producing cookies. In this example, the cost of Caroline’s factory is $30 per hour, and the cost of a worker is $10 per hour. If she hires 1 worker, her total cost is $40 per hour. If she hires 2 workers, her total cost is $50 per hour, and so on. With this information, the table now shows how the number of workers Caroline hires is related to the quantity of cookies she produces and to her total cost of production.

Our goal in the next several chapters is to study firms’ production and pricing decisions. For this purpose, the most important relationship in Table 1 is between quantity produced (in the second column) and total costs (in the sixth column). Panel (b) of Figure 2 graphs these two columns of data with the quantity produced on the horizontal axis and total cost on the vertical axis. This graph is called the total-cost curve.

Now compare the total-cost curve in panel (b) with the production function in panel (a). These two curves are opposite sides of the same coin. The total-cost curve gets steeper as the amount produced rises, whereas the production function gets flatter as production rises. These changes in slope occur for the same reason. High production of cookies means that Caroline’s kitchen is crowded with many workers. Because the kitchen is crowded, each additional worker adds less to production, reflecting diminishing marginal product. Therefore, the production function is relatively flat. But now turn this logic around: When the kitchen is crowded, producing an additional cookie requires a lot of additional labor and is thus very costly. Therefore, when the quantity produced is large, the total-cost curve is relatively steep.

**Quick Quiz** If Farmer Jones plants no seeds on his farm, he gets no harvest. If he plants 1 bag of seeds, he gets 3 bushels of wheat. If he plants 2 bags, he gets 5 bushels. If he plants 3 bags, he gets 6 bushels. A bag of seeds costs $100, and seeds are his only cost. Use these data to graph the farmer’s production function and total-cost curve. Explain their shapes.
The Various Measures of Cost

Our analysis of Caroline’s Cookie Factory demonstrated how a firm’s total cost reflects its production function. From data on a firm’s total cost, we can derive several related measures of cost, which will turn out to be useful when we analyze production and pricing decisions in future chapters. To see how these related measures are derived, we consider the example in Table 2. This table presents cost data on Caroline’s neighbor—Conrad’s Coffee Shop.

The first column of the table shows the number of cups of coffee that Conrad might produce, ranging from 0 to 10 cups per hour. The second column shows Conrad’s total cost of producing coffee. Figure 3 plots Conrad’s total-cost curve. The quantity of coffee (from the first column) is on the horizontal axis, and total cost (from the second column) is on the vertical axis. Conrad’s total-cost curve has a shape similar to Caroline’s. In particular, it becomes steeper as the quantity produced rises, which (as we have discussed) reflects diminishing marginal product.

**Fixed and Variable Costs**

Conrad’s total cost can be divided into two types. Some costs, called fixed costs, do not vary with the quantity of output produced. They are incurred even if the

---

**Table 2**

<table>
<thead>
<tr>
<th>Quantity of Coffee (cups per hour)</th>
<th>Total Cost</th>
<th>Fixed Cost</th>
<th>Variable Cost</th>
<th>Average Fixed Cost</th>
<th>Average Variable Cost</th>
<th>Average Total Cost</th>
<th>Marginal Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$3.00</td>
<td>$3.00</td>
<td>$0.00</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>$0.30</td>
</tr>
<tr>
<td>1</td>
<td>3.30</td>
<td>3.00</td>
<td>0.30</td>
<td>$3.00</td>
<td>$0.30</td>
<td>$3.30</td>
<td>0.50</td>
</tr>
<tr>
<td>2</td>
<td>3.80</td>
<td>3.00</td>
<td>0.80</td>
<td>1.50</td>
<td>0.40</td>
<td>1.90</td>
<td>0.70</td>
</tr>
<tr>
<td>3</td>
<td>4.50</td>
<td>3.00</td>
<td>1.50</td>
<td>1.00</td>
<td>0.50</td>
<td>1.50</td>
<td>0.90</td>
</tr>
<tr>
<td>4</td>
<td>5.40</td>
<td>3.00</td>
<td>2.40</td>
<td>0.75</td>
<td>0.60</td>
<td>1.35</td>
<td>1.10</td>
</tr>
<tr>
<td>5</td>
<td>6.50</td>
<td>3.00</td>
<td>3.50</td>
<td>0.60</td>
<td>0.70</td>
<td>1.30</td>
<td>1.30</td>
</tr>
<tr>
<td>6</td>
<td>7.80</td>
<td>3.00</td>
<td>4.80</td>
<td>0.50</td>
<td>0.80</td>
<td>1.30</td>
<td>1.50</td>
</tr>
<tr>
<td>7</td>
<td>9.30</td>
<td>3.00</td>
<td>6.30</td>
<td>0.43</td>
<td>0.90</td>
<td>1.33</td>
<td>1.70</td>
</tr>
<tr>
<td>8</td>
<td>11.00</td>
<td>3.00</td>
<td>8.00</td>
<td>0.38</td>
<td>1.00</td>
<td>1.38</td>
<td>1.90</td>
</tr>
<tr>
<td>9</td>
<td>12.90</td>
<td>3.00</td>
<td>9.90</td>
<td>0.33</td>
<td>1.10</td>
<td>1.43</td>
<td>2.10</td>
</tr>
<tr>
<td>10</td>
<td>15.00</td>
<td>3.00</td>
<td>12.00</td>
<td>0.30</td>
<td>1.20</td>
<td>1.50</td>
<td></td>
</tr>
</tbody>
</table>
firm produces nothing at all. Conrad’s fixed costs include any rent he pays because this cost is the same regardless of how much coffee he produces. Similarly, if Conrad needs to hire a full-time bookkeeper to pay bills, regardless of the quantity of coffee produced, the bookkeeper’s salary is a fixed cost. The third column in Table 2 shows Conrad’s fixed cost, which in this example is $3.00.

Some of the firm’s costs, called **variable costs**, change as the firm alters the quantity of output produced. Conrad’s variable costs include the cost of coffee beans, milk, sugar, and paper cups: The more cups of coffee Conrad makes, the more of these items he needs to buy. Similarly, if Conrad has to hire more workers to make more cups of coffee, the salaries of these workers are variable costs. The fourth column of the table shows Conrad’s variable cost. The variable cost is 0 if he produces nothing, $0.30 if he produces 1 cup of coffee, $0.80 if he produces 2 cups, and so on.

A firm’s total cost is the sum of fixed and variable costs. In Table 2, total cost in the second column equals fixed cost in the third column plus variable cost in the fourth column.

### Average and Marginal Cost

As the owner of his firm, Conrad has to decide how much to produce. A key part of this decision is how his costs will vary as he changes the level of production. In making this decision, Conrad might ask his production supervisor the following two questions about the cost of producing coffee:

- How much does it cost to make the typical cup of coffee?
- How much does it cost to increase production of coffee by 1 cup?
Although at first these two questions might seem to have the same answer, they do not. Both answers will turn out to be important for understanding how firms make production decisions.

To find the cost of the typical unit produced, we would divide the firm’s costs by the quantity of output it produces. For example, if the firm produces 2 cups of coffee per hour, its total cost is $3.80, and the cost of the typical cup is $3.80/2, or $1.90. Total cost divided by the quantity of output is called average total cost. Because total cost is the sum of fixed and variable costs, average total cost can be expressed as the sum of average fixed cost and average variable cost. Average fixed cost is the fixed cost divided by the quantity of output, and average variable cost is the variable cost divided by the quantity of output.

Although average total cost tells us the cost of the typical unit, it does not tell us how much total cost will change as the firm alters its level of production. The last column in Table 2 shows the amount that total cost rises when the firm increases production by 1 unit of output. This number is called marginal cost. For example, if Conrad increases production from 2 to 3 cups, total cost rises from $3.80 to $4.50, so the marginal cost of the third cup of coffee is $4.50 minus $3.80, or $0.70. In the table, the marginal cost appears halfway between two rows because it represents the change in total cost as quantity of output increases from one level to another.

It may be helpful to express these definitions mathematically:

\[
\text{Average total cost} = \frac{\text{Total cost}}{\text{Quantity}}
\]

\[
ATC = \frac{TC}{Q}
\]

and

\[
\text{Marginal cost} = \frac{\text{Change in total cost}}{\text{Change in quantity}}
\]

\[
MC = \frac{\Delta TC}{\Delta Q}.
\]

Here \(\Delta\), the Greek letter delta, represents the change in a variable. These equations show how average total cost and marginal cost are derived from total cost. Average total cost tells us the cost of a typical unit of output if total cost is divided evenly over all the units produced. Marginal cost tells us the increase in total cost that arises from producing an additional unit of output. As we will see more fully in the next chapter, business managers like Conrad need to keep in mind the concepts of average total cost and marginal cost when deciding how much of their product to supply to the market.

**Cost Curves and Their Shapes**

Just as in previous chapters we found graphs of supply and demand useful when analyzing the behavior of markets, we will find graphs of average and marginal cost useful when analyzing the behavior of firms. Figure 4 graphs Conrad’s costs using the data from Table 2. The horizontal axis measures the quantity the firm produces, and the vertical axis measures marginal and average costs. The graph shows four curves: average total cost (\(ATC\)), average fixed cost (\(AFC\)), average variable cost (\(AVC\)), and marginal cost (\(MC\)).

The cost curves shown here for Conrad’s Coffee Shop have some features that are common to the cost curves of many firms in the economy. Let’s examine...
three features in particular: the shape of the marginal-cost curve, the shape of the average-total-cost curve, and the relationship between marginal and average total cost.

**Rising Marginal Cost** Conrad’s marginal cost rises with the quantity of output produced. This reflects the property of diminishing marginal product. When Conrad produces a small quantity of coffee, he has few workers, and much of his equipment is not used. Because he can easily put these idle resources to use, the marginal product of an extra worker is large, and the marginal cost of an extra cup of coffee is small. By contrast, when Conrad produces a large quantity of coffee, his shop is crowded with workers, and most of his equipment is fully utilized. Conrad can produce more coffee by adding workers, but these new workers have to work in crowded conditions and may have to wait to use the equipment. Therefore, when the quantity of coffee produced is already high, the marginal product of an extra worker is low, and the marginal cost of an extra cup of coffee is large.

**U-Shaped Average Total Cost** Conrad’s average-total-cost curve is U-shaped, as shown in Figure 4. To understand why, remember that average total cost is the sum of average fixed cost and average variable cost. Average fixed cost always declines as output rises because the fixed cost is spread over a larger number of units. Average variable cost typically rises as output increases because of diminishing marginal product.

---

**Figure 4**

Conrad’s Average-Cost and Marginal-Cost Curves
This figure shows the average total cost (ATC), average fixed cost (AFC), average variable cost (AVC), and marginal cost (MC) for Conrad’s Coffee Shop. All of these curves are obtained by graphing the data in Table 2. These cost curves show three features that are typical of many firms: (1) Marginal cost rises with the quantity of output. (2) The average-total-cost curve is U-shaped. (3) The marginal-cost curve crosses the average-total-cost curve at the minimum of average total cost.
Average total cost reflects the shapes of both average fixed cost and average variable cost. At very low levels of output, such as 1 or 2 cups per hour, average total cost is very high. Even though average variable cost is low, average fixed cost is high because the fixed cost is spread over only a few units. As output increases, the fixed cost is spread more widely. Average fixed cost declines, rapidly at first and then more slowly. As a result, average total cost also declines until the firm’s output reaches 5 cups of coffee per hour, when average total cost is $1.30 per cup. When the firm produces more than 6 cups per hour, however, the increase in average variable cost becomes the dominant force, and average total cost starts rising. The tug of war between average fixed cost and average variable cost generates the U-shape in average total cost.

The bottom of the U-shape occurs at the quantity that minimizes average total cost. This quantity is sometimes called the efficient scale of the firm. For Conrad, the efficient scale is 5 or 6 cups of coffee per hour. If he produces more or less than this amount, his average total cost rises above the minimum of $1.30. At lower levels of output, average total cost is higher than $1.30 because the fixed cost is spread over so few units. At higher levels of output, average total cost is higher than $1.30 because the marginal product of inputs has diminished significantly. At the efficient scale, these two forces are balanced to yield the lowest average total cost.

**The Relationship between Marginal Cost and Average Total Cost**

If you look at Figure 4 (or back at Table 2), you will see something that may be surprising at first. **Whenever marginal cost is less than average total cost, average total cost is falling. Whenever marginal cost is greater than average total cost, average total cost is rising.** This feature of Conrad’s cost curves is not a coincidence from the particular numbers used in the example: It is true for all firms.

To see why, consider an analogy. Average total cost is like your cumulative grade point average. Marginal cost is like the grade in the next course you will take. If your grade in your next course is less than your grade point average, your grade point average will fall. If your grade in your next course is higher than your grade point average, your grade point average will rise. The mathematics of average and marginal costs is exactly the same as the mathematics of average and marginal grades.

This relationship between average total cost and marginal cost has an important corollary: **The marginal-cost curve crosses the average-total-cost curve at its minimum.** Why? At low levels of output, marginal cost is below average total cost, so average total cost is falling. But after the two curves cross, marginal cost rises above average total cost. For the reason we have just discussed, average total cost must start to rise at this level of output. Hence, this point of intersection is the minimum of average total cost. As you will see in the next chapter, minimum average total cost plays a key role in the analysis of competitive firms.

**Typical Cost Curves**

In the examples we have studied so far, the firms exhibit diminishing marginal product and, therefore, rising marginal cost at all levels of output. This simplifying assumption was useful because it allowed us to focus on the key features of cost curves that will prove useful in analyzing firm behavior. Yet actual firms are usually more complicated than this. In many firms, marginal product does not start...
to fall immediately after the first worker is hired. Depending on the production process, the second or third worker might have a higher marginal product than the first because a team of workers can divide tasks and work more productively than a single worker. Firms exhibiting this pattern would experience increasing marginal product for a while before diminishing marginal product set in.

Figure 5 shows the cost curves for such a firm, including average total cost ($ATC$), average fixed cost ($AFC$), average variable cost ($AVC$), and marginal cost ($MC$). At low levels of output, the firm experiences increasing marginal product, and the marginal-cost curve falls. Eventually, the firm starts to experience diminishing marginal product, and the marginal-cost curve starts to rise. This combination of increasing then diminishing marginal product also makes the average-variable-cost curve U-shaped.

Despite these differences from our previous example, the cost curves shown here share the three properties that are most important to remember:

- Marginal cost eventually rises with the quantity of output.
- The average-total-cost curve is U-shaped.
- The marginal-cost curve crosses the average-total-cost curve at the minimum of average total cost.

**Quick Quiz** Suppose Honda's total cost of producing 4 cars is $225,000 and its total cost of producing 5 cars is $250,000. What is the average total cost of producing 5 cars? What is the marginal cost of the fifth car? • Draw the marginal-cost curve and the average-total-cost curve for a typical firm, and explain why these curves cross where they do.

**Figure 5**

Cost Curves for a Typical Firm
Many firms experience increasing marginal product before diminishing marginal product. As a result, they have cost curves shaped like those in this figure. Notice that marginal cost and average variable cost fall for a while before starting to rise.
We noted earlier in this chapter that a firm’s costs might depend on the time horizon under consideration. Let’s examine more precisely why this might be the case.

**The Relationship between Short-Run and Long-Run Average Total Cost**

For many firms, the division of total costs between fixed and variable costs depends on the time horizon. Consider, for instance, a car manufacturer such as Ford Motor Company. Over a period of only a few months, Ford cannot adjust the number or sizes of its car factories. The only way it can produce additional cars is to hire more workers at the factories it already has. The cost of these factories is, therefore, a fixed cost in the short run. By contrast, over a period of several years, Ford can expand the size of its factories, build new factories, or close old ones. Thus, the cost of its factories is a variable cost in the long run.

Because many decisions are fixed in the short run but variable in the long run, a firm’s long-run cost curves differ from its short-run cost curves. Figure 6 shows an example. The figure presents three short-run average-total-cost curves—for a small, medium, and large factory. It also presents the long-run average-total-cost curve. As the firm moves along the long-run curve, it is adjusting the size of the factory to the quantity of production.

This graph shows how short-run and long-run costs are related. The long-run average-total-cost curve is a much flatter U-shape than the short-run average-total-cost curve. In addition, all the short-run curves lie on or above the long-run average-total-cost curve.
curve. These properties arise because firms have greater flexibility in the long run. In essence, in the long run, the firm gets to choose which short-run curve it wants to use. But in the short run, it has to use whatever short-run curve it has chosen in the past.

The figure shows an example of how a change in production alters costs over different time horizons. When Ford wants to increase production from 1,000 to 1,200 cars per day, it has no choice in the short run but to hire more workers at its existing medium-sized factory. Because of diminishing marginal product, average total cost rises from $10,000 to $12,000 per car. In the long run, however, Ford can expand both the size of the factory and its workforce, and average total cost returns to $10,000.

How long does it take a firm to get to the long run? The answer depends on the firm. It can take a year or longer for a major manufacturing firm, such as a car company, to build a larger factory. By contrast, a person running a coffee shop can buy another coffee maker within a few hours. There is, therefore, no single answer to how long it takes a firm to adjust its production facilities.

**ECONOMIES AND DISECONOMIES OF SCALE**

The shape of the long-run average-total-cost curve conveys important information about the production processes that a firm has available for manufacturing a good. In particular, it tells us how costs vary with the scale—that is, the size—of a firm’s operations. When long-run average total cost declines as output increases, there are said to be **economies of scale**. When long-run average total cost rises as output increases, there are said to be **diseconomies of scale**. When long-run average total cost does not vary with the level of output, there are said to be **constant returns to scale**. In this example, Ford has economies of scale at low levels of output, constant returns to scale at intermediate levels of output, and diseconomies of scale at high levels of output.

What might cause economies or diseconomies of scale? Economies of scale often arise because higher production levels allow specialization among workers, which permits each worker to become better at a specific task. For instance, if Ford hires a large number of workers and produces a large number of cars, it can reduce costs with modern assembly-line production. Diseconomies of scale can arise because of coordination problems that are inherent in any large organization. The more cars Ford produces, the more stretched the management team becomes, and the less effective the managers become at keeping costs down.

This analysis shows why long-run average-total-cost curves are often U-shaped. At low levels of production, the firm benefits from increased size because it can take advantage of greater specialization. Coordination problems, meanwhile, are not yet acute. By contrast, at high levels of production, the benefits of specialization have already been realized, and coordination problems become more severe as the firm grows larger. Thus, long-run average total cost is falling at low levels of production because of increasing specialization and rising at high levels of production because of increasing coordination problems.

**Quick Quiz** If Boeing produces 9 jets per month, its long-run total cost is $9.0 million per month. If it produces 10 jets per month, its long-run total cost is $9.5 million per month. Does Boeing exhibit economies or diseconomies of scale?
The purpose of this chapter has been to develop some tools that we can use to study how firms make production and pricing decisions. You should now understand what economists mean by the term costs and how costs vary with the quantity of output a firm produces. To refresh your memory, Table 3 summarizes some of the definitions we have encountered.

Lessons from a Pin Factory

“Jack of all trades, master of none.” This well-known adage helps explain why firms sometimes experience economies of scale. A person who tries to do everything usually ends up doing nothing very well. If a firm wants its workers to be as productive as they can be, it is often best to give each worker a limited task that he or she can master. But this is possible only if a firm employs many workers and produces a large quantity of output.

In his celebrated book *An Inquiry into the Nature and Causes of the Wealth of Nations*, Adam Smith described a visit he made to a pin factory. Smith was impressed by the specialization among the workers and the resulting economies of scale. He wrote,

*One man draws out the wire, another straightens it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations; to put it on is a peculiar business; to whiten it is another; it is even a trade by itself to put them into paper.*

Smith reported that because of this specialization, the pin factory produced thousands of pins per worker every day. He conjectured that if the workers had chosen to work separately, rather than as a team of specialists, “they certainly could not each of them make twenty, perhaps not one pin a day.” In other words, because of specialization, a large pin factory could achieve higher output per worker and lower average cost per pin than a small pin factory.

The specialization that Smith observed in the pin factory is prevalent in the modern economy. If you want to build a house, for instance, you could try to do all the work yourself. But most people turn to a builder, who in turn hires carpenters, plumbers, electricians, painters, and many other types of workers. These workers specialize in particular jobs, and this allows them to become better at their jobs than if they were generalists. Indeed, the use of specialization to achieve economies of scale is one reason modern societies are as prosperous as they are.

CONCLUSION

The purpose of this chapter has been to develop some tools that we can use to study how firms make production and pricing decisions. You should now understand what economists mean by the term *costs* and how costs vary with the quantity of output a firm produces. To refresh your memory, Table 3 summarizes some of the definitions we have encountered.

By themselves, a firm’s cost curves do not tell us what decisions the firm will make. But they are an important component of that decision, as we will begin to see in the next chapter.
TABLE 3

The Many Types of Cost: A Summary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Mathematical Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit costs</td>
<td>Costs that require an outlay of money by the firm</td>
<td></td>
</tr>
<tr>
<td>Implicit costs</td>
<td>Costs that do not require an outlay of money by the firm</td>
<td></td>
</tr>
<tr>
<td>Fixed costs</td>
<td>Costs that do not vary with the quantity of output produced</td>
<td>FC</td>
</tr>
<tr>
<td>Variable costs</td>
<td>Costs that vary with the quantity of output produced</td>
<td>VC</td>
</tr>
<tr>
<td>Total cost</td>
<td>The market value of all the inputs that a firm uses in production</td>
<td>TC = FC + VC</td>
</tr>
<tr>
<td>Average fixed cost</td>
<td>Fixed cost divided by the quantity of output</td>
<td>AFC = FC / Q</td>
</tr>
<tr>
<td>Average variable cost</td>
<td>Variable cost divided by the quantity of output</td>
<td>AVC = VC / Q</td>
</tr>
<tr>
<td>Average total cost</td>
<td>Total cost divided by the quantity of output</td>
<td>ATC = TC / Q</td>
</tr>
<tr>
<td>Marginal cost</td>
<td>The increase in total cost that arises from an extra unit of production</td>
<td>MC = ΔTC / ΔQ</td>
</tr>
</tbody>
</table>

SUMMARY

• The goal of firms is to maximize profit, which equals total revenue minus total cost.

• When analyzing a firm’s behavior, it is important to include all the opportunity costs of production. Some of the opportunity costs, such as the wages a firm pays its workers, are explicit. Other opportunity costs, such as the wages the firm owner gives up by working in the firm rather than taking another job, are implicit.

• A firm’s costs reflect its production process. A typical firm’s production function gets flatter as the quantity of an input increases, displaying the property of diminishing marginal product. As a result, a firm’s total-cost curve gets steeper as the quantity produced rises.

• A firm’s total costs can be divided between fixed costs and variable costs. Fixed costs are costs that do not change when the firm alters the quantity of output produced. Variable costs are costs that change when the firm alters the quantity of output produced.

• From a firm’s total cost, two related measures of cost are derived. Average total cost is total cost divided by the quantity of output. Marginal cost is the amount by which total cost rises if output increases by 1 unit.
• When analyzing firm behavior, it is often useful to graph average total cost and marginal cost. For a typical firm, marginal cost rises with the quantity of output. Average total cost first falls as output increases and then rises as output increases further. The marginal-cost curve always crosses the average-total-cost curve at the minimum of average total cost.

• A firm’s costs often depend on the time horizon considered. In particular, many costs are fixed in the short run but variable in the long run. As a result, when the firm changes its level of production, average total cost may rise more in the short run than in the long run.

**KEY CONCEPTS**

- total revenue, p. 244
- total cost, p. 244
- profit, p. 244
- explicit costs, p. 245
- implicit costs, p. 245
- economic profit, p. 246
- accounting profit, p. 246
- production function, p. 247
- marginal product, p. 248
- diminishing marginal product, p. 249
- fixed costs, p. 250
- variable costs, p. 251
- average total cost, p. 252
- average fixed cost, p. 252
- average variable cost, p. 252
- marginal cost, p. 252
- efficient scale, p. 254
- economies of scale, p. 257
- diseconomies of scale, p. 257
- constant returns to scale, p. 257

**QUESTIONS FOR REVIEW**

1. What is the relationship between a firm’s total revenue, profit, and total cost?
2. Give an example of an opportunity cost that an accountant might not count as a cost. Why would the accountant ignore this cost?
3. What is marginal product, and what does it mean if it is diminishing?
4. Draw a production function that exhibits diminishing marginal product of labor. Draw the associated total-cost curve. (In both cases, be sure to label the axes.) Explain the shapes of the two curves you have drawn.
5. Define total cost, average total cost, and marginal cost. How are they related?
6. Draw the marginal-cost and average-total-cost curves for a typical firm. Explain why the curves have the shapes that they do and why they cross where they do.
7. How and why does a firm’s average-total-cost curve differ in the short run and in the long run?
8. Define economies of scale and explain why they might arise. Define diseconomies of scale and explain why they might arise.
1. This chapter discusses many types of costs: opportunity cost, total cost, fixed cost, variable cost, average total cost, and marginal cost. Fill in the type of cost that best completes each sentence:
   a. What you give up for taking some action is called the _____.
   b. _____ is falling when marginal cost is below it and rising when marginal cost is above it.
   c. A cost that does not depend on the quantity produced is a _____.
   d. In the ice-cream industry in the short run, _____ includes the cost of cream and sugar but not the cost of the factory.
   e. Profits equal total revenue less _____.
   f. The cost of producing an extra unit of output is the _____.

2. Your aunt is thinking about opening a hardware store. She estimates that it would cost $500,000 per year to rent the location and buy the stock. In addition, she would have to quit her $50,000 per year job as an accountant.
   a. Define opportunity cost.
   b. What is your aunt’s opportunity cost of running a hardware store for a year? If your aunt thought she could sell $510,000 worth of merchandise in a year, should she open the store? Explain.

3. A commercial fisherman notices the following relationship between hours spent fishing and the quantity of fish caught:

<table>
<thead>
<tr>
<th>Hours (in pounds)</th>
<th>Quantity of Fish (in pounds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 hours</td>
<td>0 lb.</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
</tr>
</tbody>
</table>

   a. What is the marginal product of each hour spent fishing?
   b. Use these data to graph the fisherman’s production function. Explain its shape.
   c. The fisherman has a fixed cost of $10 (his pole). The opportunity cost of his time is $5 per hour. Graph the fisherman’s total-cost curve. Explain its shape.

4. Nimbus, Inc., makes brooms and then sells them door-to-door. Here is the relationship between the number of workers and Nimbus’s output in a given day:

<table>
<thead>
<tr>
<th>Workers</th>
<th>Output</th>
<th>Marginal Product</th>
<th>Total Cost</th>
<th>Average Total Cost</th>
<th>Marginal Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>120</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>155</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

   a. Fill in the column of marginal products. What pattern do you see? How might you explain it?
   b. A worker costs $100 a day, and the firm has fixed costs of $200. Use this information to fill in the column for total cost.
   c. Fill in the column for average total cost. (Recall that \( ATC = TC/Q \)) What pattern do you see?
d. Now fill in the column for marginal cost. (Recall that $MC = \Delta TC/\Delta Q$.) What pattern do you see?

e. Compare the column for marginal product and the column for marginal cost. Explain the relationship.

f. Compare the column for average total cost and the column for marginal cost. Explain the relationship.

5. You are the Chief Financial Officer for a firm that sells digital music players. Your firm has the following average total cost schedule:

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Average Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>600 players</td>
<td>$300</td>
</tr>
<tr>
<td>601</td>
<td>$301</td>
</tr>
</tbody>
</table>

Your current level of production is 600 devices, all of which have been sold. Someone calls, desperate to buy one of your music players. The caller offers you $550 for it. Should you accept the offer? Why or why not?

6. Consider the following cost information for a pizzeria:

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Total Cost</th>
<th>Variable Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 dozen pizzas</td>
<td>$300</td>
<td>$0</td>
</tr>
<tr>
<td>1</td>
<td>350</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>390</td>
<td>90</td>
</tr>
<tr>
<td>3</td>
<td>420</td>
<td>120</td>
</tr>
<tr>
<td>4</td>
<td>450</td>
<td>150</td>
</tr>
<tr>
<td>5</td>
<td>490</td>
<td>190</td>
</tr>
<tr>
<td>6</td>
<td>540</td>
<td>240</td>
</tr>
</tbody>
</table>

a. What is the pizzeria’s fixed cost?

b. Construct a table in which you calculate the marginal cost per dozen pizzas using the information on total cost. Also, calculate the marginal cost per dozen pizzas using the information on variable cost. What is the relationship between these sets of numbers? Comment.

7. You are thinking about setting up a lemonade stand. The stand itself costs $200. The ingredients for each cup of lemonade cost $0.50.

a. What is your fixed cost of doing business?

b. What is your variable cost per cup?

b. Construct a table showing your total cost, average total cost, and marginal cost for output levels varying from 0 to 10 gallons. (Hint: There are 16 cups in a gallon.) Draw the three cost curves.

8. Your cousin Vinnie owns a painting company with fixed costs of $200 and the following schedule for variable costs:

<table>
<thead>
<tr>
<th>Quantity of Houses Painted per Month</th>
<th>Variable Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$10</td>
</tr>
<tr>
<td>2</td>
<td>$20</td>
</tr>
<tr>
<td>3</td>
<td>$40</td>
</tr>
<tr>
<td>4</td>
<td>$80</td>
</tr>
<tr>
<td>5</td>
<td>$160</td>
</tr>
<tr>
<td>6</td>
<td>$320</td>
</tr>
<tr>
<td>7</td>
<td>$640</td>
</tr>
</tbody>
</table>

Calculate average fixed cost, average variable cost, and average total cost for each quantity. What is the efficient scale of the painting company?

9. The city government is considering two tax proposals:
   - A lump-sum tax of $300 on each producer of hamburgers.
   - A tax of $1 per burger, paid by producers of hamburgers.

a. Which of the following curves—average fixed cost, average variable cost, average total cost, and marginal cost—would shift as a result of the lump-sum tax? Why? Show this in a graph. Label the graph as precisely as possible.
b. Which of these same four curves would shift as a result of the per-burger tax? Why? Show this in a new graph. Label the graph as precisely as possible.

10. Jane’s Juice Bar has the following cost schedules:

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Variable Cost</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 vats of juice</td>
<td>$0</td>
<td>$30</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>70</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>130</td>
</tr>
<tr>
<td>6</td>
<td>135</td>
<td>165</td>
</tr>
</tbody>
</table>

a. Calculate average variable cost, average total cost, and marginal cost for each quantity.
b. Graph all three curves. What is the relationship between the marginal-cost curve and the average-total-cost curve? Between the marginal-cost curve and the average-variable-cost curve? Explain.

11. A firm has fixed cost of $100 and average variable cost of $5 \times Q$, where $Q$ is the number of units produced.

a. Construct a table showing total cost for $Q$ from 0 to 10.
b. Graph the firm’s curves for marginal cost and average total cost.
c. How does marginal cost change with $Q$? What does this suggest about the firm’s production process?

12. Consider the following table of long-run total costs for three different firms:

<table>
<thead>
<tr>
<th>Quantity</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm A</td>
<td>$60</td>
<td>$70</td>
<td>$80</td>
<td>$90</td>
<td>$100</td>
<td>$110</td>
<td>$120</td>
</tr>
<tr>
<td>Firm B</td>
<td>11</td>
<td>24</td>
<td>39</td>
<td>56</td>
<td>75</td>
<td>96</td>
<td>119</td>
</tr>
<tr>
<td>Firm C</td>
<td>21</td>
<td>34</td>
<td>49</td>
<td>66</td>
<td>85</td>
<td>106</td>
<td>129</td>
</tr>
</tbody>
</table>

Does each of these firms experience economies of scale or diseconomies of scale?
Firms in Competitive Markets

If your local gas station raised its price for gasoline by 20 percent, it would see a large drop in the amount of gasoline it sold. Its customers would quickly switch to buying their gasoline at other gas stations. By contrast, if your local water company raised the price of water by 20 percent, it would see only a small decrease in the amount of water it sold. People might water their lawns less often and buy more water-efficient showerheads, but they would be hard-pressed to reduce water consumption greatly and would be unlikely to find another supplier. The difference between the gasoline market and the water market is obvious: Many firms supply gasoline to the local market, but only one firm supplies water. As you might expect, this difference in market structure shapes the pricing and production decisions of the firms that operate in these markets.

In this chapter, we examine the behavior of competitive firms, such as your local gas station. You may recall that a market is competitive if each buyer and seller is small compared to the size of the market and, therefore, has little ability to influence market prices. By contrast, if a firm can influence the market price of the good it sells, it is said to have market power. Later in the book, we examine the behavior of firms with market power, such as your local water company.
Our analysis of competitive firms in this chapter sheds light on the decisions that lie behind the supply curve in a competitive market. Not surprisingly, we will find that a market supply curve is tightly linked to firms’ costs of production. Less obvious, however, is the question of which among a firm’s many types of cost—fixed, variable, average, and marginal—are most relevant for its supply decisions. We will see that all these measures of cost play important and interrelated roles.

### WHAT IS A COMPETITIVE MARKET?

Our goal in this chapter is to examine how firms make production decisions in competitive markets. As a background for this analysis, we begin by reviewing what a competitive market is.

#### The Meaning of Competition

A **competitive market**, sometimes called a *perfectly competitive market*, has two characteristics:

- There are many buyers and many sellers in the market.
- The goods offered by the various sellers are largely the same.

As a result of these conditions, the actions of any single buyer or seller in the market have a negligible impact on the market price. Each buyer and seller takes the market price as given.

As an example, consider the market for milk. No single consumer of milk can influence the price of milk because each buyer purchases a small amount relative to the size of the market. Similarly, each dairy farmer has limited control over the price because many other sellers are offering milk that is essentially identical. Because each seller can sell all he wants at the going price, he has little reason to charge less, and if he charges more, buyers will go elsewhere. Buyers and sellers in competitive markets must accept the price the market determines and, therefore, are said to be *price takers*.

In addition to the foregoing two conditions for competition, there is a third condition sometimes thought to characterize perfectly competitive markets:

- Firms can freely enter or exit the market.

If, for instance, anyone can decide to start a dairy farm, and if any existing dairy farmer can decide to leave the dairy business, then the dairy industry would satisfy this condition. Much of the analysis of competitive firms does not need the assumption of free entry and exit because this condition is not necessary for firms to be price takers. Yet, as we will see later in this chapter, if there is free entry and exit in a competitive market, it is a powerful force shaping the long-run equilibrium.

#### The Revenue of a Competitive Firm

A firm in a competitive market, like most other firms in the economy, tries to maximize profit (total revenue minus total cost). To see how it does this, we first consider the revenue of a competitive firm. To keep matters concrete, let’s consider a specific firm: the Vaca Family Dairy Farm.
The Vaca Farm produces a quantity of milk, $Q$, and sells each unit at the market price, $P$. The farm’s total revenue is $P \times Q$. For example, if a gallon of milk sells for $6 and the farm sells 1,000 gallons, its total revenue is $6,000.

Because the Vaca Farm is small compared to the world market for milk, it takes the price as given by market conditions. This means, in particular, that the price of milk does not depend on the number of gallons that the Vaca Farm produces and sells. If the Vacas double the amount of milk they produce to 2,000 gallons, the price of milk remains the same, and their total revenue doubles to $12,000. As a result, total revenue is proportional to the amount of output.

Table 1 shows the revenue for the Vaca Family Dairy Farm. The first two columns show the amount of output the farm produces and the price at which it sells its output. The third column is the farm’s total revenue. The table assumes that the price of milk is $6 a gallon, so total revenue is $6 times the number of gallons.

Just as the concepts of average and marginal were useful in the preceding chapter when analyzing costs, they are also useful when analyzing revenue. To see what these concepts tell us, consider these two questions:

- How much revenue does the farm receive for the typical gallon of milk?
- How much additional revenue does the farm receive if it increases production of milk by 1 gallon?

The last two columns in Table 1 answer these questions.

The fourth column in the table shows average revenue, which is total revenue (from the third column) divided by the amount of output (from the first column). Average revenue tells us how much revenue a firm receives for the typical unit sold. In Table 1, you can see that average revenue equals $6, the price of a gallon of milk. This illustrates a general lesson that applies not only to competitive firms but to other firms as well. Average revenue is total revenue ($P \times Q$) divided by the quantity ($Q$). Therefore, for all firms, average revenue equals the price of the good.

<table>
<thead>
<tr>
<th>Quantity ($Q$)</th>
<th>Price ($P$)</th>
<th>Total Revenue ($TR = P \times Q$)</th>
<th>Average Revenue ($AR = TR / Q$)</th>
<th>Marginal Revenue ($MR = \Delta TR / \Delta Q$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 gallon</td>
<td>$6</td>
<td>$6</td>
<td>$6</td>
<td>$6</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>$6</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>18</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>24</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>30</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>36</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>42</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>48</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

**TABLE 1**

Total, Average, and Marginal Revenue for a Competitive Firm
The fifth column shows **marginal revenue**, which is the change in total revenue from the sale of each additional unit of output. In Table 1, marginal revenue equals $6, the price of a gallon of milk. This result illustrates a lesson that applies only to competitive firms. Total revenue is $P \times Q$, and $P$ is fixed for a competitive firm. Therefore, when $Q$ rises by 1 unit, total revenue rises by $P$ dollars. **For competitive firms, marginal revenue equals the price of the good.**

**Quick Quiz** When a competitive firm doubles the amount it sells, what happens to the price of its output and its total revenue?

---

**PROFIT MAXIMIZATION AND THE COMPETITIVE FIRM’S SUPPLY CURVE**

The goal of a competitive firm is to maximize profit, which equals total revenue minus total cost. We have just discussed the firm’s revenue, and in the preceding chapter, we discussed the firm’s costs. We are now ready to examine how a competitive firm maximizes profit and how that decision determines its supply curve.

**A Simple Example of Profit Maximization**

Let’s begin our analysis of the firm’s supply decision with the example in Table 2. In the first column of the table is the number of gallons of milk the Vaca Family Dairy Farm produces. The second column shows the farm’s total revenue, which is $6 times the number of gallons. The third column shows the farm’s total cost. Total cost includes fixed costs, which are $3 in this example, and variable costs, which depend on the quantity produced.

The fourth column shows the farm’s profit, which is computed by subtracting total cost from total revenue. If the farm produces nothing, it has a loss of $3 (its fixed cost). If it produces 1 gallon, it has a profit of $1. If it produces 2 gallons, it has a profit of $4 and so on. Because the Vaca family’s goal is to maximize profit, it chooses to produce the quantity of milk that makes profit as large as possible. In this example, profit is maximized when the farm produces 4 or 5 gallons of milk, for a profit of $7.

There is another way to look at the Vaca Farm’s decision: The Vacas can find the profit-maximizing quantity by comparing the marginal revenue and marginal cost from each unit produced. The fifth and sixth columns in Table 2 compute marginal revenue and marginal cost from the changes in total revenue and total cost, and the last column shows the change in profit for each additional gallon produced. The first gallon of milk the farm produces has a marginal revenue of $6 and a marginal cost of $2; hence, producing that gallon increases profit by $4 (from $3 to $1). The second gallon produced has a marginal revenue of $6 and a marginal cost of $3, so that gallon increases profit by $3 (from $1 to $4). As long as marginal revenue exceeds marginal cost, increasing the quantity produced raises profit. Once the Vaca Farm has reached 5 gallons of milk, however, the situation changes. The sixth gallon would have a marginal revenue of $6 and a marginal cost of $7, so producing it would reduce profit by $1 (from $7 to $6). As a result, the Vacas would not produce beyond 5 gallons.

One of the *Ten Principles of Economics* in Chapter 1 is that rational people think at the margin. We now see how the Vaca Family Dairy Farm can apply this
principle. If marginal revenue is greater than marginal cost—as it is at 1, 2, or 3 gallons—the Vacas should increase the production of milk because it will put more money in their pockets (marginal revenue) than it takes out (marginal cost). If marginal revenue is less than marginal cost—as it is at 6, 7, or 8 gallons—the Vacas should decrease production. If the Vacas think at the margin and make incremental adjustments to the level of production, they are naturally led to produce the profit-maximizing quantity.

THE MARGINAL-COST CURVE AND THE FIRM’S SUPPLY DECISION

To extend this analysis of profit maximization, consider the cost curves in Figure 1. These cost curves have the three features that, as we discussed in the previous chapter, are thought to describe most firms: The marginal-cost curve ($MC$) is upward sloping. The average-total-cost curve ($ATC$) is U-shaped. And the marginal-cost curve crosses the average-total-cost curve at the minimum of average total cost. The figure also shows a horizontal line at the market price ($P$). The price line is horizontal because the firm is a price taker: The price of the firm’s output is the same regardless of the quantity that the firm decides to produce. Keep in mind that, for a competitive firm, the firm’s price equals both its average revenue ($AR$) and its marginal revenue ($MR$).

We can use Figure 1 to find the quantity of output that maximizes profit. Imagine that the firm is producing at $Q_1$. At this level of output, marginal revenue is greater than marginal cost. That is, if the firm raised its level of production and sales by 1 unit, the additional revenue ($MR_1$) would exceed the additional cost ($MC_1$). Profit, which equals total revenue minus total cost, would increase.
Hence, if marginal revenue is greater than marginal cost, as it is at $Q_1$, the firm can increase profit by increasing production.

A similar argument applies when output is at $Q_2$. In this case, marginal cost is greater than marginal revenue. If the firm reduced production by 1 unit, the costs saved ($MC_2$) would exceed the revenue lost ($MR_2$). Therefore, if marginal revenue is less than marginal cost, as it is at $Q_2$, the firm can increase profit by reducing production.

Where do these marginal adjustments to production end? Regardless of whether the firm begins with production at a low level (such as $Q_1$) or at a high level (such as $Q_2$), the firm will eventually adjust production until the quantity produced reaches $Q_{MAX}$. This analysis yields three general rules for profit maximization:

- If marginal revenue is greater than marginal cost, the firm should increase its output.
- If marginal cost is greater than marginal revenue, the firm should decrease its output.
- At the profit-maximizing level of output, marginal revenue and marginal cost are exactly equal.

These rules are the key to rational decision making by a profit-maximizing firm. They apply not only to competitive firms but, as we will see in the next chapter, to other types of firms as well.
We can now see how the competitive firm decides the quantity of its good to supply to the market. Because a competitive firm is a price taker, its marginal revenue equals the market price. For any given price, the competitive firm’s profit-maximizing quantity of output is found by looking at the intersection of the price with the marginal-cost curve. In Figure 1, that quantity of output is $Q_{\text{MAX}}$.

Suppose that the price prevailing in this market rises, perhaps because of an increase in market demand. Figure 2 shows how a competitive firm responds to the price increase. When the price is $P_1$, the firm produces quantity $Q_1$, the quantity that equates marginal cost to the price. When the price rises to $P_2$, the firm finds that marginal revenue is now higher than marginal cost at the previous level of output, so the firm increases production. The new profit-maximizing quantity is $Q_2$, at which marginal cost equals the new higher price. In essence, because the firm’s marginal-cost curve determines the quantity of the good the firm is willing to supply at any price, the marginal-cost curve is also the competitive firm’s supply curve. There are, however, some caveats to that conclusion, which we examine next.

**The Firm’s Short-Run Decision to Shut Down**

So far, we have been analyzing the question of how much a competitive firm will produce. In certain circumstances, however, the firm will decide to shut down and not produce anything at all.

Here we should distinguish between a temporary shutdown of a firm and the permanent exit of a firm from the market. A **shutdown** refers to a short-run decision not to produce anything during a specific period of time because of current market conditions. **Exit** refers to a long-run decision to leave the market. The short-run and long-run decisions differ because most firms cannot avoid their fixed costs in the short run but can do so in the long run. That is, a firm that shuts down temporarily still has to pay its fixed costs, whereas a firm that exits the market does not have to pay any costs at all, fixed or variable.

![Marginal Cost as the Competitive Firm’s Supply Curve](image)
For example, consider the production decision that a farmer faces. The cost of the land is one of the farmer’s fixed costs. If the farmer decides not to produce any crops one season, the land lies fallow, and he cannot recover this cost. When making the short-run decision whether to shut down for a season, the fixed cost of land is said to be a *sunk cost*. By contrast, if the farmer decides to leave farming altogether, he can sell the land. When making the long-run decision whether to exit the market, the cost of land is not sunk. (We return to the issue of sunk costs shortly.)

Now let’s consider what determines a firm’s shutdown decision. If the firm shuts down, it loses all revenue from the sale of its product. At the same time, it saves the variable costs of making its product (but must still pay the fixed costs). Thus, *the firm shuts down if the revenue that it would get from producing is less than its variable costs of production.*

A small bit of mathematics can make this shutdown criterion more useful. If $TR$ stands for total revenue and $VC$ stands for variable costs, then the firm’s decision can be written as

$$\text{Shut down if } TR < VC.$$

The firm shuts down if total revenue is less than variable cost. By dividing both sides of this inequality by the quantity $Q$, we can write it as

$$\text{Shut down if } TR/Q < VC/Q.$$

The left side of the inequality, $TR/Q$, is total revenue $P \times Q$ divided by quantity $Q$, which is average revenue, most simply expressed as the good’s price, $P$. The right side of the inequality, $VC/Q$, is average variable cost, $AVC$. Therefore, the firm’s shutdown criterion can be restated as

$$\text{Shut down if } P < AVC.$$

That is, a firm chooses to shut down if the price of the good is less than the average variable cost of production. This criterion is intuitive: When choosing to produce, the firm compares the price it receives for the typical unit to the average variable cost that it must incur to produce the typical unit. If the price doesn’t cover the average variable cost, the firm is better off stopping production altogether. The firm will be losing money (since it still has to pay fixed costs), but it would lose even more money staying open. The firm can reopen in the future if conditions change so that price exceeds average variable cost.

We now have a full description of a competitive firm’s profit-maximizing strategy. If the firm produces anything, it produces the quantity at which marginal cost equals the price of the good. Yet if the price is less than average variable cost at that quantity, the firm is better off shutting down and not producing anything. These results are illustrated in Figure 3. *The competitive firm’s short-run supply curve is the portion of its marginal-cost curve that lies above average variable cost.*

**Spilt Milk and Other Sunk Costs**

Sometime in your life you may have been told, “Don’t cry over spilt milk,” or “Let bygones be bygones.” These adages hold a deep truth about rational decision making. Economists say that a cost is a *sunk cost* when it has already been com-
mitted and cannot be recovered. Because nothing can be done about sunk costs, you can ignore them when making decisions about various aspects of life, including business strategy.

Our analysis of the firm’s shutdown decision is one example of the irrelevance of sunk costs. We assume that the firm cannot recover its fixed costs by temporarily stopping production. That is, regardless of the quantity of output supplied, and even if it is zero, the firm still has to pay its fixed costs. As a result, the fixed costs are sunk in the short run, and the firm can ignore them when deciding how much to produce. The firm’s short-run supply curve is the part of the marginal-cost curve that lies above average variable cost, and the size of the fixed cost does not matter for this supply decision.

The irrelevance of sunk costs is also important when making personal decisions. Imagine, for instance, that you place a $15 value on seeing a newly released movie. You buy a ticket for $10, but before entering the theater, you lose the ticket. Should you buy another ticket? Or should you now go home and refuse to pay a total of $20 to see the movie? The answer is that you should buy another ticket. The benefit of seeing the movie ($15) still exceeds the opportunity cost (the $10 for the second ticket). The $10 you paid for the lost ticket is a sunk cost. As with spilt milk, there is no point in crying about it.

NEAR-EMPTY RESTAURANTS AND OFF-SEASON MINIATURE GOLF

Have you ever walked into a restaurant for lunch and found it almost empty? Why, you might have asked, does the restaurant even bother to stay open? It might seem that the revenue from the few customers could not possibly cover the cost of running the restaurant.
In making the decision whether to open for lunch, a restaurant owner must keep in mind the distinction between fixed and variable costs. Many of a restaurant’s costs—the rent, kitchen equipment, tables, plates, silverware, and so on—are fixed. Shutting down during lunch would not reduce these costs. In other words, these costs are sunk in the short run. When the owner is deciding whether to serve lunch, only the variable costs—the price of the additional food and the wages of the extra staff—are relevant. The owner shuts down the restaurant at lunchtime only if the revenue from the few lunchtime customers fails to cover the restaurant’s variable costs.

An operator of a miniature-golf course in a summer resort community faces a similar decision. Because revenue varies substantially from season to season, the firm must decide when to open and when to close. Once again, the fixed costs—the costs of buying the land and building the course—are irrelevant in making this decision. The miniature-golf course should be open for business only during those times of year when its revenue exceeds its variable costs.

**The Firm’s Long-Run Decision to Exit or Enter a Market**

A firm’s long-run decision to exit a market is similar to its shutdown decision. If the firm exits, it will again lose all revenue from the sale of its product, but now it will save not only its variable costs of production but also its fixed costs. Thus, the firm exits the market if the revenue it would get from producing is less than its total costs.

We can again make this criterion more useful by writing it mathematically. If \( TR \) stands for total revenue, and \( TC \) stands for total cost, then the firm’s exit criterion can be written as

\[
\text{Exit if } TR < TC.
\]

The firm exits if total revenue is less than total cost. By dividing both sides of this inequality by quantity \( Q \), we can write it as

\[
\text{Exit if } TR / Q < TC / Q.
\]

We can simplify this further by noting that \( TR / Q \) is average revenue, which equals the price \( P \), and that \( TC / Q \) is average total cost, \( ATC \). Therefore, the firm’s exit criterion is

\[
\text{Exit if } P < ATC.
\]

That is, a firm chooses to exit if the price of its good is less than the average total cost of production.

A parallel analysis applies to an entrepreneur who is considering starting a firm. The firm will enter the market if such an action would be profitable, which occurs if the price of the good exceeds the average total cost of production. The entry criterion is

\[
\text{Enter if } P > ATC.
\]

The criterion for entry is exactly the opposite of the criterion for exit.
We can now describe a competitive firm’s long-run profit-maximizing strategy. If the firm is in the market, it produces the quantity at which marginal cost equals the price of the good. Yet if the price is less than average total cost at that quantity, the firm chooses to exit (or not enter) the market. These results are illustrated in Figure 4. The competitive firm’s long-run supply curve is the portion of its marginal-cost curve that lies above average total cost.

**Measuring Profit in Our Graph for the Competitive Firm**

As we analyze exit and entry, it is useful to be able to analyze the firm’s profit in more detail. Recall that profit equals total revenue \((TR)\) minus total cost \((TC)\):

\[
\text{Profit} = TR - TC.
\]

We can rewrite this definition by multiplying and dividing the right side by \(Q\):

\[
\text{Profit} = (TR/Q - TC/Q) \times Q.
\]

But note that \(TR/Q\) is average revenue, which is the price, \(P\), and \(TC/Q\) is average total cost, \(ATC\). Therefore,

\[
\text{Profit} = (P - ATC) \times Q.
\]

This way of expressing the firm’s profit allows us to measure profit in our graphs.

Panel (a) of Figure 5 shows a firm earning positive profit. As we have already discussed, the firm maximizes profit by producing the quantity at which price equals marginal cost. Now look at the shaded rectangle. The height of the rectangle

---

**Figure 4**

**The Competitive Firm’s Long-Run Supply Curve**

In the long run, the competitive firm’s supply curve is its marginal-cost curve \((MC)\) above average total cost \((ATC)\). If the price falls below average total cost, the firm is better off exiting the market.
is \( P - ATC \), the difference between price and average total cost. The width of the rectangle is \( Q \), the quantity produced. Therefore, the area of the rectangle is \((P - ATC) \times Q\), which is the firm’s profit.

Similarly, panel (b) of this figure shows a firm with losses (negative profit). In this case, maximizing profit means minimizing losses, a task accomplished once again by producing the quantity at which price equals marginal cost. Now consider the shaded rectangle. The height of the rectangle is \( ATC - P \), and the width is \( Q \). The area is \((ATC - P) \times Q\), which is the firm’s loss. Because a firm in this situation is not making enough revenue to cover its average total cost, the firm would choose in the long run to exit the market.

**QUICK QUIZ** How does a competitive firm determine its profit-maximizing level of output? Explain. • When does a profit-maximizing competitive firm decide to shut down? When does a profit-maximizing competitive firm decide to exit a market?

**THE SUPPLY CURVE IN A COMPETITIVE MARKET**

Now that we have examined the supply decision of a single firm, we can discuss the supply curve for a market. There are two cases to consider. First, we examine a market with a fixed number of firms. Second, we examine a market in which the number of firms can change as old firms exit the market and new firms enter. Both cases are important, for each applies over a specific time horizon. Over short
periods of time, it is often difficult for firms to enter and exit, so the assumption of a fixed number of firms is appropriate. But over long periods of time, the number of firms can adjust to changing market conditions.

**The Short Run: Market Supply with a Fixed Number of Firms**

Consider first a market with 1,000 identical firms. For any given price, each firm supplies a quantity of output so that its marginal cost equals the price, as shown in panel (a) of Figure 6. That is, as long as price is above average variable cost, each firm’s marginal-cost curve is its supply curve. The quantity of output supplied to the market equals the sum of the quantities supplied by each of the 1,000 individual firms. Thus, to derive the market supply curve, we add the quantity supplied by each firm in the market. As panel (b) of Figure 6 shows, because the firms are identical, the quantity supplied to the market is 1,000 times the quantity supplied by each firm.

**The Long Run: Market Supply with Entry and Exit**

Now consider what happens if firms are able to enter or exit the market. Let’s suppose that everyone has access to the same technology for producing the good and access to the same markets to buy the inputs into production. Therefore, all firms and all potential firms have the same cost curves.

In the short run, the number of firms in the market is fixed. As a result, the market supply curve, shown in panel (b), reflects the individual firms’ marginal-cost curves, shown in panel (a). Here, in a market of 1,000 firms, the quantity of output supplied to the market is 1,000 times the quantity supplied by each firm.
Decisions about entry and exit in a market of this type depend on the incentives facing the owners of existing firms and the entrepreneurs who could start new firms. If firms already in the market are profitable, then new firms will have an incentive to enter the market. This entry will expand the number of firms, increase the quantity of the good supplied, and drive down prices and profits. Conversely, if firms in the market are making losses, then some existing firms will exit the market. Their exit will reduce the number of firms, decrease the quantity of the good supplied, and drive up prices and profits. At the end of this process of entry and exit, firms that remain in the market must be making zero economic profit.

Recall that we can write a firm’s profit as

\[ \text{Profit} = (P - ATC) \times Q. \]

This equation shows that an operating firm has zero profit if and only if the price of the good equals the average total cost of producing that good. If price is above average total cost, profit is positive, which encourages new firms to enter. If price is less than average total cost, profit is negative, which encourages some firms to exit. The process of entry and exit ends only when price and average total cost are driven to equality.

This analysis has a surprising implication. We noted earlier in the chapter that competitive firms maximize profits by choosing a quantity at which price equals marginal cost. We just noted that free entry and exit force price to equal average total cost. But if price is to equal both marginal cost and average total cost, these two measures of cost must equal each other. Marginal cost and average total cost are equal, however, only when the firm is operating at the minimum of average total cost. Recall from the preceding chapter that the level of production with lowest average total cost is called the firm’s efficient scale. Therefore, in the long-run equilibrium of a competitive market with free entry and exit, firms must be operating at their efficient scale.

Panel (a) of Figure 7 shows a firm in such a long-run equilibrium. In this figure, price \( P \) equals marginal cost \( MC \), so the firm is profit-maximizing. Price also equals average total cost \( ATC \), so profits are zero. New firms have no incentive to enter the market, and existing firms have no incentive to leave the market.

From this analysis of firm behavior, we can determine the long-run supply curve for the market. In a market with free entry and exit, there is only one price consistent with zero profit—the minimum of average total cost. As a result, the long-run market supply curve must be horizontal at this price, as illustrated by the perfectly elastic supply curve in panel (b) of Figure 7. Any price above this level would generate profit, leading to entry and an increase in the total quantity supplied. Any price below this level would generate losses, leading to exit and a decrease in the total quantity supplied. Eventually, the number of firms in the market adjusts so that price equals the minimum of average total cost, and there are enough firms to satisfy all the demand at this price.

**Why Do Competitive Firms Stay in Business If They Make Zero Profit?**

At first, it might seem odd that competitive firms earn zero profit in the long run. After all, people start businesses to make a profit. If entry eventually drives profit to zero, there might seem to be little reason to stay in business.
To understand the zero-profit condition more fully, recall that profit equals total revenue minus total cost and that total cost includes all the opportunity costs of the firm. In particular, total cost includes the time and money that the firm owners devote to the business. In the zero-profit equilibrium, the firm’s revenue must compensate the owners for these opportunity costs.

Consider an example. Suppose that, to start his farm, a farmer had to invest $1 million, which otherwise he could have deposited in a bank to earn $50,000 a year in interest. In addition, he had to give up another job that would have paid him $30,000 a year. Then the farmer’s opportunity cost of farming includes both the interest he could have earned and the forgone wages—a total of $80,000. Even if his profit is driven to zero, his revenue from farming compensates him for these opportunity costs.

Keep in mind that accountants and economists measure costs differently. As we discussed in the previous chapter, accountants keep track of explicit costs but not implicit costs. That is, they measure costs that require an outflow of money from the firm, but they do not include the opportunity costs of production that do not involve an outflow of money. As a result, in the zero-profit equilibrium, economic profit is zero, but accounting profit is positive. Our farmer’s accountant, for instance, would conclude that the farmer earned an accounting profit of $80,000, which is enough to keep the farmer in business.

A Shift in Demand in the Short Run and Long Run

Now that we have a more complete understanding of how firms make supply decisions, we can better explain how markets respond to changes in demand.
Because firms can enter and exit in the long run but not in the short run, the response of a market to a change in demand depends on the time horizon. To see this, let’s trace the effects of a shift in demand over time.

Suppose the market for milk begins in a long-run equilibrium. Firms are earning zero profit, so price equals the minimum of average total cost. Panel (a) of Figure 8 shows the situation. The long-run equilibrium is point A, the quantity sold in the market is $Q_1$, and the price is $P_1$.

Now suppose scientists discover that milk has miraculous health benefits. As a result, the demand curve for milk shifts outward from $D_1$ to $D_2$, as in panel (b). The short-run equilibrium moves from point A to point B; as a result, the quantity rises from $Q_1$ to $Q_2$, and the price rises from $P_1$ to $P_2$. All of the existing firms respond to the higher price by raising the amount produced. Because each firm’s supply curve reflects its marginal-cost curve, how much they each increase production is determined by the marginal-cost curve. In the new short-run equilibrium, the price of milk exceeds average total cost, so the firms are making positive profit.

Over time, the profit in this market encourages new firms to enter. Some farmers may switch to milk from other farm products, for example. As the number of firms grows, the short-run supply curve shifts to the right from $S_1$ to $S_2$, as in panel (c), and this shift causes the price of milk to fall. Eventually, the price is driven back down to the minimum of average total cost, profits are zero, and firms stop entering. Thus, the market reaches a new long-run equilibrium, point C. The price of milk has returned to $P_1$, but the quantity produced has risen to $Q_3$. Each firm is again producing at its efficient scale, but because more firms are in the dairy business, the quantity of milk produced and sold is higher.

**WHY THE LONG-RUN SUPPLY CURVE MIGHT SLOPE UPWARD**

So far, we have seen that entry and exit can cause the long-run market supply curve to be perfectly elastic. The essence of our analysis is that there are a large number of potential entrants, each of which faces the same costs. As a result, the long-run market supply curve is horizontal at the minimum of average total cost. When the demand for the good increases, the long-run result is an increase in the number of firms and in the total quantity supplied, without any change in the price.

There are, however, two reasons that the long-run market supply curve might slope upward. The first is that some resource used in production may be available only in limited quantities. For example, consider the market for farm products. Anyone can choose to buy land and start a farm, but the quantity of land is limited. As more people become farmers, the price of farmland is bid up, which raises the costs of all farmers in the market. Thus, an increase in demand for farm products cannot induce an increase in quantity supplied without also inducing a rise in farmers’ costs, which in turn means a rise in price. The result is a long-run market supply curve that is upward sloping, even with free entry into farming.

A second reason for an upward-sloping supply curve is that firms may have different costs. For example, consider the market for painters. Anyone can enter the market for painting services, but not everyone has the same costs. Costs vary in part because some people work faster than others and in part because some people have better alternative uses of their time than others. For any given price, those with lower costs are more likely to enter than those with higher costs. To increase the quantity of painting services supplied, additional entrants must be encouraged to enter the market. Because these new entrants have higher costs,
The market starts in a long-run equilibrium, shown as point A in panel (a). In this equilibrium, each firm makes zero profit, and the price equals the minimum average total cost. Panel (b) shows what happens in the short run when demand rises from $D_1$ to $D_2$. The equilibrium goes from point A to point B, price rises from $P_1$ to $P_2$, and the quantity sold in the market rises from $Q_1$ to $Q_2$. Because price now exceeds average total cost, firms make profits, which over time encourage new firms to enter the market. This entry shifts the short-run supply curve to the right from $S_1$ to $S_2$, as shown in panel (c). In the new long-run equilibrium, point C, price has returned to $P_1$, but the quantity sold has increased to $Q_3$. Profits are again zero, price is back to the minimum of average total cost, but the market has more firms to satisfy the greater demand.

(a) Initial Condition

(b) Short-Run Response

(c) Long-Run Response
the price must rise to make entry profitable for them. Thus, the long-run market supply curve for painting services slopes upward even with free entry into the market.

Notice that if firms have different costs, some firms earn profit even in the long run. In this case, the price in the market reflects the average total cost of the marginal firm—the firm that would exit the market if the price were any lower. This firm earns zero profit, but firms with lower costs earn positive profit. Entry does not eliminate this profit because would-be entrants have higher costs than firms already in the market. Higher-cost firms will enter only if the price rises, making the market profitable for them.

Thus, for these two reasons, a higher price may be necessary to induce a larger quantity supplied, in which case the long-run supply curve is upward sloping rather than horizontal. Nonetheless, the basic lesson about entry and exit remains true. Because firms can enter and exit more easily in the long run than in the short run, the long-run supply curve is typically more elastic than the short-run supply curve.

**Quick Quiz** In the long run with free entry and exit, is the price in a market equal to marginal cost, average total cost, both, or neither? Explain with a diagram.

**Conclusion: Behind the Supply Curve**

We have been discussing the behavior of profit-maximizing firms that supply goods in perfectly competitive markets. You may recall from Chapter 1 that one of the Ten Principles of Economics is that rational people think at the margin. This chapter has applied this idea to the competitive firm. Marginal analysis has given us a theory of the supply curve in a competitive market and, as a result, a deeper understanding of market outcomes.

We have learned that when you buy a good from a firm in a competitive market, you can be assured that the price you pay is close to the cost of producing that good. In particular, if firms are competitive and profit maximizing, the price of a good equals the marginal cost of making that good. In addition, if firms can freely enter and exit the market, the price also equals the lowest possible average total cost of production.

Although we have assumed throughout this chapter that firms are price takers, many of the tools developed here are also useful for studying firms in less competitive markets. We now turn to an examination of the behavior of firms with market power. Marginal analysis will again be useful, but it will have quite different implications.
SUMMARY

- Because a competitive firm is a price taker, its revenue is proportional to the amount of output it produces. The price of the good equals both the firm’s average revenue and its marginal revenue.

- To maximize profit, a firm chooses a quantity of output such that marginal revenue equals marginal cost. Because marginal revenue for a competitive firm equals the market price, the firm chooses quantity so that price equals marginal cost. Thus, the firm’s marginal-cost curve is its supply curve.

- In the short run when a firm cannot recover its fixed costs, the firm will choose to shut down temporarily if the price of the good is less than average variable cost. In the long run when the firm can recover both fixed and variable costs, it will choose to exit if the price is less than average total cost.

- In a market with free entry and exit, profits are driven to zero in the long run. In this long-run equilibrium, all firms produce at the efficient scale, price equals the minimum of average total cost, and the number of firms adjusts to satisfy the quantity demanded at this price.

- Changes in demand have different effects over different time horizons. In the short run, an increase in demand raises prices and leads to profits, and a decrease in demand lowers prices and leads to losses. But if firms can freely enter and exit the market, then in the long run, the number of firms adjusts to drive the market back to the zero-profit equilibrium.

KEY CONCEPTS

- competitive market, p. 266
- marginal revenue, p. 268
- sunk cost, p. 272
- average revenue, p. 267

QUESTIONS FOR REVIEW

1. What is meant by a competitive firm?
2. Explain the difference between a firm’s revenue and its profit. Which do firms maximize?
3. Draw the cost curves for a typical firm. For a given price, explain how the firm chooses the level of output that maximizes profit. At that level of output, show on your graph the total revenue of the firm. Show its total costs.
4. Under what conditions will a firm shut down temporarily? Explain.
5. Under what conditions will a firm exit a market? Explain.
6. Does a firm’s price equal marginal cost in the short run, in the long run, or both? Explain.
7. Does a firm’s price equal the minimum of average total cost in the short run, in the long run, or both? Explain.
8. Are market supply curves typically more elastic in the short run or in the long run? Explain.
5. Ball Bearings Inc. faces costs of production as follows:

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Fixed Costs</th>
<th>Total Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$100</td>
<td>$100</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>150</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>170</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>240</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>360</td>
</tr>
</tbody>
</table>

a. Calculate the company’s average fixed costs, average variable costs, average total costs, and marginal costs.
b. The price of a case of ball bearings is $50. Seeing that she can’t make a profit, the Chief Executive Officer (CEO) decides to shut down operations. What are the firm’s profits/losses? Was this a wise decision? Explain.
c. Vaguely remembering his introductory economics course, the Chief Financial Officer tells the CEO it is better to produce 1 case of ball bearings, because marginal revenue equals marginal cost at that quantity. What are the firm’s profits/losses at that level of production? Was this the best decision? Explain.

a. Draw a diagram describing the typical firm in the industry.
b. Hi-Tech Printing Company invents a new process that sharply reduces the cost of printing books. What happens to Hi-Tech’s profits and the price of books in the short run when Hi-Tech’s patent prevents other firms from using the new technology?
c. What happens in the long run when the patent expires and other firms are free to use the technology?
7. A firm in a competitive market receives $500 in total revenue and has marginal revenue of $10. What is the average revenue, and how many units were sold?

8. A profit-maximizing firm in a competitive market is currently producing 100 units of output. It has average revenue of $10, average total cost of $8, and fixed costs of $200.
   a. What is profit?
   b. What is marginal cost?
   c. What is average variable cost?
   d. Is the efficient scale of the firm more than, less than, or exactly 100 units?

9. The market for fertilizer is perfectly competitive. Firms in the market are producing output, but are currently making economic losses.
   a. How does the price of fertilizer compare to the average total cost, the average variable cost, and the marginal cost of producing fertilizer?
   b. Draw two graphs, side by side, illustrating the present situation for the typical firm and in the market.
   c. Assuming there is no change in demand or the firms’ cost curves, explain what will happen in the long run to the price of fertilizer, marginal cost, average total cost, the quantity supplied by each firm, and the total quantity supplied to the market.

10. Suppose that the U.S. textile industry is competitive, and there is no international trade in textiles. In long-run equilibrium, the price per unit of cloth is $30.
   a. Describe the equilibrium using graphs for the entire market and for an individual producer.
   b. Assuming that textile producers in other countries are willing to sell large quantities of cloth in the United States for only $25 per unit.
   c. What is the short-run effect of these imports on the quantity produced by an individual producer? What is the short-run effect on profits? Illustrate your answer with a graph.
   d. What is the long-run effect on the number of U.S. firms in the industry?

11. An industry currently has 100 firms, all of which have fixed costs of $16 and average variable cost as follows:

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Average Variable Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

   a. Compute marginal cost and average total cost.
   b. The price is currently $10. What is the total quantity supplied in the market?
   c. As this market makes the transition to its long-run equilibrium, will the price rise or fall? Will the quantity demanded rise or fall? Will the quantity supplied by each firm rise or fall?
   d. Graph the long-run supply curve for this market.

12. Suppose there are 1,000 hot pretzel stands operating in New York City. Each stand has the usual U-shaped average-total-cost curve. The market demand curve for pretzels slopes downward, and the market for pretzels is in long-run competitive equilibrium.
   a. Draw the current equilibrium, using graphs for the entire market and for an individual stand.
   b. The city decides to restrict the number of pretzel-stand licenses, reducing the number of stands to only 800. What effect will this action have on the market and on an individual stand that is still operating? Draw graphs to illustrate your answer.
   c. Suppose that the city decides to charge a fee for the 800 licenses, all of which are quickly sold. How will the size of the fee affect the number of pretzels sold by an individual stand? How will it affect the price of pretzels in the city?
   d. The city wants to raise as much revenue as possible, while ensuring that all 800 licenses are sold. How high should the city set the license fee? Show the answer on your graph.
13. Assume that the gold-mining industry is competitive.
   a. Illustrate a long-run equilibrium using diagrams for the gold market and for a representative gold mine.
   b. Suppose that an increase in jewelry demand induces a surge in the demand for gold. Using your diagrams from part (a), show what happens in the short run to the gold market and to each existing gold mine.
   c. If the demand for gold remains high, what would happen to the price over time? Specifically, would the new long-run equilibrium price be above, below, or equal to the short-run equilibrium price in part (b)? Is it possible for the new long-run equilibrium price to be above the original long-run equilibrium price? Explain.

14. Analyze the two following situations for firms in competitive markets:
   a. Suppose that \( TC = 100 + 15q \), where \( TC \) is total cost and \( q \) is the quantity produced. What is the minimum price necessary for this firm to produce any output in the short run?
   b. Suppose that \( MC = 4q \), where \( MC \) is marginal cost. The perfectly competitive firm maximizes profits by producing 10 units of output. At what price does it sell these units?
Monopoly

If you own a personal computer, it probably uses some version of Windows, the operating system sold by the Microsoft Corporation. When Microsoft first designed Windows many years ago, it applied for and received a copyright from the government. The copyright gives Microsoft the exclusive right to make and sell copies of the Windows operating system. If a person wants to buy a copy of Windows, he or she has little choice but to give Microsoft the approximately $100 that the firm has decided to charge for its product. Microsoft is said to have a monopoly in the market for Windows.

Microsoft’s business decisions are not well described by the model of firm behavior we developed in the previous chapter. In that chapter, we analyzed competitive markets, in which there are many firms offering essentially identical products, so each firm has little influence over the price it receives. By contrast, a monopoly such as Microsoft has no close competitors and, therefore, has the power to influence the market price of its product. While a competitive firm is a price taker, a monopoly firm is a price maker.

In this chapter, we examine the implications of this market power. We will see that market power alters the relationship between a firm’s costs and the price at which it sells its product. A competitive firm takes the price of its output as given by the market and then chooses the quantity it will supply so that price equals marginal cost. By contrast, a monopoly charges a price that exceeds marginal cost.
This result is clearly true in the case of Microsoft’s Windows. The marginal cost of Windows—the extra cost that Microsoft incurs by printing one more copy of the program onto a CD—is only a few dollars. The market price of Windows is many times marginal cost.

It is not surprising that monopolies charge high prices for their products. Customers of monopolies might seem to have little choice but to pay whatever the monopoly charges. But if so, why does a copy of Windows not cost $1,000? Or $10,000? The reason is that if Microsoft sets the price that high, fewer people would buy the product. People would buy fewer computers, switch to other operating systems, or make illegal copies. A monopoly firm can control the price of the good it sells, but because a high price reduces the quantity that its customers buy, the monopoly’s profits are not unlimited.

As we examine the production and pricing decisions of monopolies, we also consider the implications of monopoly for society as a whole. Monopoly firms, like competitive firms, aim to maximize profit. But this goal has very different ramifications for competitive and monopoly firms. In competitive markets, self-interested consumers and producers behave as if they are guided by an invisible hand to promote general economic well-being. By contrast, because monopoly firms are unchecked by competition, the outcome in a market with a monopoly is often not in the best interest of society.

One of the Ten Principles of Economics in Chapter 1 is that governments can sometimes improve market outcomes. The analysis in this chapter sheds more light on this principle. As we examine the problems that monopolies raise for society, we discuss the various ways in which government policymakers might respond to these problems. The U.S. government, for example, keeps a close eye on Microsoft’s business decisions. In 1994, it blocked Microsoft from buying Intuit, a leading seller of personal finance software, on the grounds that combining the two firms would concentrate too much market power. Similarly, in 1998, the U.S. Department of Justice objected when Microsoft started integrating its Internet browser into its Windows operating system, claiming that this addition would extend the firm’s market power into new areas. To this day, Microsoft continues to wrangle with antitrust regulators in the United States and abroad.

### WHY MONOPOLIES ARISE

A firm is a **monopoly** if it is the sole seller of its product and if its product does not have close substitutes. The fundamental cause of monopoly is **barriers to entry**: A monopoly remains the only seller in its market because other firms cannot enter the market and compete with it. Barriers to entry, in turn, have three main sources:

- **Monopoly resources**: A key resource required for production is owned by a single firm.
- **Government regulation**: The government gives a single firm the exclusive right to produce some good or service.
- **The production process**: A single firm can produce output at a lower cost than can a larger number of producers.

Let’s briefly discuss each of these.
**Monopoly Resources**

The simplest way for a monopoly to arise is for a single firm to own a key resource. For example, consider the market for water in a small town in the Old West. If dozens of town residents have working wells, the competitive model discussed in the preceding chapter describes the behavior of sellers. As a result of the competition among water suppliers, the price of a gallon is driven to equal the marginal cost of pumping an extra gallon. But if there is only one well in town and it is impossible to get water from anywhere else, then the owner of the well has a monopoly on water. Not surprisingly, the monopolist has much greater market power than any single firm in a competitive market. In the case of a necessity like water, the monopolist could command quite a high price, even if the marginal cost of pumping an extra gallon is low.

A classic example of market power arising from the ownership of a key resource is DeBeers, the South African diamond company. Founded in 1888 by Cecil Rhodes, an English businessman (and benefactor for the Rhodes scholarship), DeBeers has at times controlled up to 80 percent of the production from the world’s diamond mines. Because its market share is less than 100 percent, DeBeers is not exactly a monopoly, but the company has nonetheless exerted substantial influence over the market price of diamonds.

Although exclusive ownership of a key resource is a potential cause of monopoly, in practice monopolies rarely arise for this reason. Economies are large, and resources are owned by many people. Indeed, because many goods are traded internationally, the natural scope of their markets is often worldwide. There are, therefore, few examples of firms that own a resource for which there are no close substitutes.

**Government-Created Monopolies**

In many cases, monopolies arise because the government has given one person or firm the exclusive right to sell some good or service. Sometimes the monopoly arises from the sheer political clout of the would-be monopolist. Kings, for example, once granted exclusive business licenses to their friends and allies. At other times, the government grants a monopoly because doing so is viewed to be in the public interest.

The patent and copyright laws are two important examples. When a pharmaceutical company discovers a new drug, it can apply to the government for a patent. If the government deems the drug to be truly original, it approves the patent, which gives the company the exclusive right to manufacture and sell the drug for 20 years. Similarly, when a novelist finishes a book, she can copyright it. The copyright is a government guarantee that no one can print and sell the work without the author’s permission. The copyright makes the novelist a monopolist in the sale of her novel.

The effects of patent and copyright laws are easy to see. Because these laws give one producer a monopoly, they lead to higher prices than would occur under competition. But by allowing these monopoly producers to charge higher prices and earn higher profits, the laws also encourage some desirable behavior. Drug companies are allowed to be monopolists in the drugs they discover to encourage research. Authors are allowed to be monopolists in the sale of their books to encourage them to write more and better books.
Thus, the laws governing patents and copyrights have benefits and costs. The benefits of the patent and copyright laws are the increased incentives for creative activity. These benefits are offset, to some extent, by the costs of monopoly pricing, which we examine fully later in this chapter.

**Natural Monopolies**

An industry is a natural monopoly when a single firm can supply a good or service to an entire market at a lower cost than could two or more firms. A natural monopoly arises when there are economies of scale over the relevant range of output. Figure 1 shows the average total costs of a firm with economies of scale. In this case, a single firm can produce any amount of output at least cost. That is, for any given amount of output, a larger number of firms leads to less output per firm and higher average total cost.

An example of a natural monopoly is the distribution of water. To provide water to residents of a town, a firm must build a network of pipes throughout the town. If two or more firms were to compete in the provision of this service, each firm would have to pay the fixed cost of building a network. Thus, the average total cost of water is lowest if a single firm serves the entire market.

We saw other examples of natural monopolies when we discussed public goods and common resources in Chapter 11. We noted in passing that some goods are excludable but not rival in consumption. An example is a bridge used so infrequently that it is never congested. The bridge is excludable because a toll collector can prevent someone from using it. The bridge is not rival in consumption because use of the bridge by one person does not diminish the ability of others to use it. Because there is a fixed cost of building the bridge and a negligible marginal cost of additional users, the average total cost of a trip across the bridge (the total cost divided by the number of trips) falls as the number of trips rises. Hence, the bridge is a natural monopoly.

**Figure 1**

**Economies of Scale as a Cause of Monopoly**

When a firm’s average-total-cost curve continually declines, the firm has what is called a natural monopoly. In this case, when production is divided among more firms, each firm produces less, and average total cost rises. As a result, a single firm can produce any given amount at the smallest cost.
When a firm is a natural monopoly, it is less concerned about new entrants eroding its monopoly power. Normally, a firm has trouble maintaining a monopoly position without ownership of a key resource or protection from the government. The monopolist’s profit attracts entrants into the market, and these entrants make the market more competitive. By contrast, entering a market in which another firm has a natural monopoly is unattractive. Would-be entrants know that they cannot achieve the same low costs that the monopolist enjoys because, after entry, each firm would have a smaller piece of the market.

In some cases, the size of the market is one determinant of whether an industry is a natural monopoly. Again, consider a bridge across a river. When the population is small, the bridge may be a natural monopoly. A single bridge can satisfy the entire demand for trips across the river at lowest cost. Yet as the population grows and the bridge becomes congested, satisfying the entire demand may require two or more bridges across the same river. Thus, as a market expands, a natural monopoly can evolve into a more competitive market.

**Quick Quiz** What are the three reasons that a market might have a monopoly? • Give two examples of monopolies and explain the reason for each.

---

**How Monopolies Make Production and Pricing Decisions**

Now that we know how monopolies arise, we can consider how a monopoly firm decides how much of its product to make and what price to charge for it. The analysis of monopoly behavior in this section is the starting point for evaluating whether monopolies are desirable and what policies the government might pursue in monopoly markets.

**Monopoly versus Competition**

The key difference between a competitive firm and a monopoly is the monopoly’s ability to influence the price of its output. A competitive firm is small relative to the market in which it operates and, therefore, has no power to influence the price of its output. It takes the price as given by market conditions. By contrast, because a monopoly is the sole producer in its market, it can alter the price of its good by adjusting the quantity it supplies to the market.

One way to view this difference between a competitive firm and a monopoly is to consider the demand curve that each firm faces. When we analyzed profit maximization by competitive firms in the preceding chapter, we drew the market price as a horizontal line. Because a competitive firm can sell as much or as little as it wants at this price, the competitive firm faces a horizontal demand curve, as in panel (a) of Figure 2. In effect, because the competitive firm sells a product with many perfect substitutes (the products of all the other firms in its market), the demand curve that any one firm faces is perfectly elastic.

By contrast, because a monopoly is the sole producer in its market, its demand curve is the market demand curve. Thus, the monopolist’s demand curve slopes downward for all the usual reasons, as in panel (b) of Figure 2. If the monopolist raises the price of its good, consumers buy less of it. Looked at another way, if the
monopolist reduces the quantity of output it produces and sells, the price of its output increases.

The market demand curve provides a constraint on a monopoly's ability to profit from its market power. A monopolist would prefer, if it were possible, to charge a high price and sell a large quantity at that high price. The market demand curve makes that outcome impossible. In particular, the market demand curve describes the combinations of price and quantity that are available to a monopoly firm. By adjusting the quantity produced (or equivalently, the price charged), the monopolist can choose any point on the demand curve, but it cannot choose a point off the demand curve.

What price and quantity of output will the monopolist choose? As with competitive firms, we assume that the monopolist's goal is to maximize profit. Because the firm's profit is total revenue minus total costs, our next task in explaining monopoly behavior is to examine a monopolist's revenue.

A MONOPOLY'S REVENUE

Consider a town with a single producer of water. Table 1 shows how the monopolist's revenue might depend on the amount of water produced.

The first two columns show the monopolist's demand schedule. If the monopolist produces 1 gallon of water, it can sell that gallon for $10. If it produces 2 gallons, it must lower the price to $9 to sell both gallons. If it produces 3 gallons, it must lower the price to $8. And so on. If you graphed these two columns of numbers, you would get a typical downward-sloping demand curve.
The third column of the table presents the monopolist’s total revenue. It equals the quantity sold (from the first column) times the price (from the second column). The fourth column computes the firm’s average revenue, the amount of revenue the firm receives per unit sold. We compute average revenue by taking the number for total revenue in the third column and dividing it by the quantity of output in the first column. As we discussed in the previous chapter, average revenue always equals the price of the good. This is true for monopolists as well as for competitive firms.

The last column of Table 1 computes the firm’s marginal revenue, the amount of revenue that the firm receives for each additional unit of output. We compute marginal revenue by taking the change in total revenue when output increases by 1 unit. For example, when the firm is producing 3 gallons of water, it receives total revenue of $24. Raising production to 4 gallons increases total revenue to $28. Thus, marginal revenue from the sale of the fourth gallon is $28 minus $24, or $4.

Table 1 shows a result that is important for understanding monopoly behavior: A monopolist’s marginal revenue is always less than the price of its good. For example, if the firm raises production of water from 3 to 4 gallons, it will increase total revenue by only $4, even though it will be able to sell each gallon for $7. For a monopoly, marginal revenue is lower than price because a monopoly faces a downward-sloping demand curve. To increase the amount sold, a monopoly firm must lower the price it charges to all customers. Hence, to sell the fourth gallon of water, the monopolist will get $1 less revenue for each of the first three gallons. This $3 loss accounts for the difference between the price of the fourth gallon ($7) and the marginal revenue of that fourth gallon ($4).
Marginal revenue for monopolies is very different from marginal revenue for competitive firms. When a monopoly increases the amount it sells, this action has two effects on total revenue ($P \times Q$):

- The output effect: More output is sold, so $Q$ is higher, which tends to increase total revenue.
- The price effect: The price falls, so $P$ is lower, which tends to decrease total revenue.

Because a competitive firm can sell all it wants at the market price, there is no price effect. When it increases production by 1 unit, it receives the market price for that unit, and it does not receive any less for the units it was already selling. That is, because the competitive firm is a price taker, its marginal revenue equals the price of its good. By contrast, when a monopoly increases production by 1 unit, it must reduce the price it charges for every unit it sells, and this cut in price reduces revenue on the units it was already selling. As a result, a monopoly’s marginal revenue is less than its price.

Figure 3 graphs the demand curve and the marginal-revenue curve for a monopoly firm. (Because the firm’s price equals its average revenue, the demand curve is also the average-revenue curve.) These two curves always start at the same point on the vertical axis because the marginal revenue of the first unit sold equals the price of the good. But for the reason we just discussed, the monopolist’s marginal revenue on all units after the first is less than the price of the good. Thus, a monopoly’s marginal-revenue curve lies below its demand curve.

You can see in the figure (as well as in Table 1) that marginal revenue can even become negative. Marginal revenue is negative when the price effect on revenue is greater than the output effect. In this case, when the firm produces an extra unit of output, the price falls by enough to cause the firm’s total revenue to decline, even though the firm is selling more units.
**Profit Maximization**

Now that we have considered the revenue of a monopoly firm, we are ready to examine how such a firm maximizes profit. Recall from Chapter 1 that one of the *Ten Principles of Economics* is that rational people think at the margin. This lesson is as true for monopolists as it is for competitive firms. Here we apply the logic of marginal analysis to the monopolist’s decision about how much to produce.

Figure 4 graphs the demand curve, the marginal-revenue curve, and the cost curves for a monopoly firm. All these curves should seem familiar: The demand and marginal-revenue curves are like those in Figure 3, and the cost curves are like those we encountered in the last two chapters. These curves contain all the information we need to determine the level of output that a profit-maximizing monopolist will choose.

Suppose, first, that the firm is producing at a low level of output, such as $Q_1$. In this case, marginal cost is less than marginal revenue. If the firm increased production by 1 unit, the additional revenue would exceed the additional costs, and profit would rise. Thus, when marginal cost is less than marginal revenue, the firm can increase profit by producing more units.

A similar argument applies at high levels of output, such as $Q_2$. In this case, marginal cost is greater than marginal revenue. If the firm reduced production by 1 unit, the costs saved would exceed the revenue lost. Thus, if marginal cost is greater than marginal revenue, the firm can raise profit by reducing production.

In the end, the firm adjusts its level of production until the quantity reaches $Q_{\text{MAX}}$ at which marginal revenue equals marginal cost. Thus, the monopolist’s profit-maximizing quantity of output is determined by the intersection of the marginal-revenue curve and the marginal-cost curve. In Figure 4, this intersection occurs at point A.

You might recall from the previous chapter that competitive firms also choose the quantity of output at which marginal revenue equals marginal cost. In

---

**Figure 4**

**Profit Maximization for a Monopoly**

A monopoly maximizes profit by choosing the quantity at which marginal revenue equals marginal cost (point A). It then uses the demand curve to find the price that will induce consumers to buy that quantity (point B).
following this rule for profit maximization, competitive firms and monopolies are alike. But there is also an important difference between these types of firms: The marginal revenue of a competitive firm equals its price, whereas the marginal revenue of a monopoly is less than its price. That is,

\[
\text{For a competitive firm: } P = MR = MC. \\
\text{For a monopoly firm: } P > MR = MC.
\]

The equality of marginal revenue and marginal cost at the profit-maximizing quantity is the same for both types of firms. What differs is the relationship of the price to marginal revenue and marginal cost.

How does the monopoly find the profit-maximizing price for its product? The demand curve answers this question because the demand curve relates the amount that customers are willing to pay to the quantity sold. Thus, after the monopoly firm chooses the quantity of output that equates marginal revenue and marginal cost, it uses the demand curve to find the highest price it can charge and sell that quantity. In Figure 4, the profit-maximizing price is found at point B.

We can now see a key difference between markets with competitive firms and markets with a monopoly firm: In competitive markets, price equals marginal cost. In monopolized markets, price exceeds marginal cost. As we will see in a moment, this finding is crucial to understanding the social cost of monopoly.

**A Monopoly’s Profit**

How much profit does a monopoly make? To see a monopoly firm’s profit in a graph, recall that profit equals total revenue (TR) minus total costs (TC):

\[
\text{Profit} = TR - TC.
\]

---

**FYI Why a Monopoly Does Not Have a Supply Curve**

You may have noticed that we have analyzed the price in a monopoly market using the market demand curve and the firm’s cost curves. We have not made any mention of the market supply curve. By contrast, when we analyzed prices in competitive markets beginning in Chapter 4, the two most important words were always supply and demand.

What happened to the supply curve? Although monopoly firms make decisions about what quantity to supply (in the way described in this chapter), a monopoly does not have a supply curve. A supply curve tells us the quantity that firms choose to supply at any given price. This concept makes sense when we are analyzing competitive firms, which are price takers. But a monopoly firm is a price maker, not a price taker. It is not meaningful to ask what such a firm would produce at any price because the firm sets the price at the same time it chooses the quantity to supply.

Indeed, the monopolist’s decision about how much to supply is impossible to separate from the demand curve it faces. The shape of the demand curve determines the shape of the marginal-revenue curve, which in turn determines the monopolist’s profit-maximizing quantity. In a competitive market, supply decisions can be analyzed without knowing the demand curve, but that is not true in a monopoly market. Therefore, we never talk about a monopoly’s supply curve.
We can rewrite this as

\[ \text{Profit} = (TR/Q - TC/Q) \times Q. \]

*TR/Q* is average revenue, which equals the price, *P*, and *TC/Q* is average total cost, *ATC*. Therefore,

\[ \text{Profit} = (P - ATC) \times Q. \]

This equation for profit (which also holds for competitive firms) allows us to measure the monopolist’s profit in our graph.

Consider the shaded box in Figure 5. The height of the box (the segment BC) is price minus average total cost, \( P - ATC \), which is the profit on the typical unit sold. The width of the box (the segment DC) is the quantity sold, \( Q_{\text{MAX}} \). Therefore, the area of this box is the monopoly firm’s total profit.

**MONOPOLY DRUGS VERSUS GENERIC DRUGS**

According to our analysis, prices are determined differently in monopolized markets and competitive markets. A natural place to test this theory is the market for pharmaceutical drugs because this market takes on both market structures. When a firm discovers a new drug, patent laws give the firm a monopoly on the sale of that drug. But eventually, the firm’s patent runs out, and any company can make and sell the drug. At that time, the market switches from being monopolistic to being competitive.

What should happen to the price of a drug when the patent runs out? Figure 6 shows the market for a typical drug. In this figure, the marginal cost of producing...
the drug is constant. (This is approximately true for many drugs.) During the life of the patent, the monopoly firm maximizes profit by producing the quantity at which marginal revenue equals marginal cost and charging a price well above marginal cost. But when the patent runs out, the profit from making the drug should encourage new firms to enter the market. As the market becomes more competitive, the price should fall to equal marginal cost.

Experience is, in fact, consistent with our theory. When the patent on a drug expires, other companies quickly enter and begin selling so-called generic products that are chemically identical to the former monopolist’s brand-name product. And just as our analysis predicts, the price of the competitively produced generic drug is well below the price that the monopolist was charging.

The expiration of a patent, however, does not cause the monopolist to lose all its market power. Some consumers remain loyal to the brand-name drug, perhaps out of fear that the new generic drugs are not actually the same as the drug they have been using for years. As a result, the former monopolist can continue to charge a price at least somewhat above the price charged by its new competitors.

**Quick Quiz** Explain how a monopolist chooses the quantity of output to produce and the price to charge.

**The Welfare Cost of Monopolies**

Is monopoly a good way to organize a market? We have seen that a monopoly, in contrast to a competitive firm, charges a price above marginal cost. From the standpoint of consumers, this high price makes monopoly undesirable. At the same time, however, the monopoly is earning profit from charging this high price.
From the standpoint of the owners of the firm, the high price makes monopoly very desirable. Is it possible that the benefits to the firm’s owners exceed the costs imposed on consumers, making monopoly desirable from the standpoint of society as a whole?

We can answer this question using the tools of welfare economics. Recall from Chapter 7 that total surplus measures the economic well-being of buyers and sellers in a market. Total surplus is the sum of consumer surplus and producer surplus. Consumer surplus is consumers’ willingness to pay for a good minus the amount they actually pay for it. Producer surplus is the amount producers receive for a good minus their costs of producing it. In this case, there is a single producer—the monopolist.

You can probably guess the result of this analysis. In Chapter 7, we concluded that the equilibrium of supply and demand in a competitive market is not only a natural outcome but also a desirable one. The invisible hand of the market leads to an allocation of resources that makes total surplus as large as it can be. Because a monopoly leads to an allocation of resources different from that in a competitive market, the outcome must, in some way, fail to maximize total economic well-being.

**The Deadweight Loss**

We begin by considering what the monopoly firm would do if it were run by a benevolent social planner. The social planner cares not only about the profit earned by the firm’s owners but also about the benefits received by the firm’s consumers. The planner tries to maximize total surplus, which equals producer surplus (profit) plus consumer surplus. Keep in mind that total surplus equals the value of the good to consumers minus the costs of making the good incurred by the monopoly producer.

Figure 7 analyzes how a benevolent social planner would choose the monopoly’s level of output. The demand curve reflects the value of the good to consumers, as measured by their willingness to pay for it. The marginal-cost curve reflects the costs of the monopolist. Thus, the socially efficient quantity is found where the demand curve and the marginal-cost curve intersect. Below this quantity, the value of an extra unit to consumers exceeds the cost of providing it, so increasing output would raise total surplus. Above this quantity, the cost of producing an extra unit exceeds the value of that unit to consumers, so decreasing output would raise total surplus. At the optimal quantity, the value of an extra unit to consumers exactly equals the marginal cost of production.

If the social planner were running the monopoly, the firm could achieve this efficient outcome by charging the price found at the intersection of the demand and marginal-cost curves. Thus, like a competitive firm and unlike a profit-maximizing monopoly, a social planner would charge a price equal to marginal cost. Because this price would give consumers an accurate signal about the cost of producing the good, consumers would buy the efficient quantity.

We can evaluate the welfare effects of monopoly by comparing the level of output that the monopolist chooses to the level of output that a social planner would choose. As we have seen, the monopolist chooses to produce and sell the quantity of output at which the marginal-revenue and marginal-cost curves intersect; the social planner would choose the quantity at which the demand and marginal-cost curves intersect. Figure 8 shows the comparison. The monopolist produces less than the socially efficient quantity of output.
We can also view the inefficiency of monopoly in terms of the monopolist’s price. Because the market demand curve describes a negative relationship between the price and quantity of the good, a quantity that is inefficiently low is equivalent to a price that is inefficiently high. When a monopolist charges a price above marginal cost, some potential consumers value the good at more than its marginal cost but less than the monopolist’s price. These consumers do not buy the good. Because the value these consumers place on the good is greater than the cost of providing it to them, this result is inefficient. Thus, monopoly pricing prevents some mutually beneficial trades from taking place.

The inefficiency of monopoly can be measured with a deadweight loss triangle, as illustrated in Figure 8. Because the demand curve reflects the value to consumers and the marginal-cost curve reflects the costs to the monopoly producer, the area of the deadweight loss triangle between the demand curve and the marginal-cost curve equals the total surplus lost because of monopoly pricing. It is the reduction in economic well-being that results from the monopoly’s use of its market power.

The deadweight loss caused by monopoly is similar to the deadweight loss caused by a tax. Indeed, a monopolist is like a private tax collector. As we saw in Chapter 8, a tax on a good places a wedge between consumers’ willingness to pay (as reflected in the demand curve) and producers’ costs (as reflected in the supply curve). Because a monopoly exerts its market power by charging a price above marginal cost, it places a similar wedge. In both cases, the wedge causes the quantity sold to fall short of the social optimum. The difference between the two

---

**Figure**

**The Efficient Level of Output**

A benevolent social planner who wanted to maximize total surplus in the market would choose the level of output where the demand curve and marginal-cost curve intersect. Below this level, the value of the good to the marginal buyer (as reflected in the demand curve) exceeds the marginal cost of making the good. Above this level, the value to the marginal buyer is less than marginal cost.
cases is that the government gets the revenue from a tax, whereas a private firm gets the monopoly profit.

**THE MONOPOLY’S PROFIT: A SOCIAL COST?**

It is tempting to decry monopolies for “profiteering” at the expense of the public. And indeed, a monopoly firm does earn a higher profit by virtue of its market power. According to the economic analysis of monopoly, however, the firm’s profit is not in itself necessarily a problem for society.

Welfare in a monopolized market, like all markets, includes the welfare of both consumers and producers. Whenever a consumer pays an extra dollar to a producer because of a monopoly price, the consumer is worse off by a dollar, and the producer is better off by the same amount. This transfer from the consumers of the good to the owners of the monopoly does not affect the market’s total surplus—the sum of consumer and producer surplus. In other words, the monopoly profit itself represents not a reduction in the size of the economic pie but merely a bigger slice for producers and a smaller slice for consumers. Unless consumers are for some reason more deserving than producers—a normative judgment about equity that goes beyond the realm of economic efficiency—the monopoly profit is not a social problem.

The problem in a monopolized market arises because the firm produces and sells a quantity of output below the level that maximizes total surplus. The deadweight loss measures how much the economic pie shrinks as a result. This inefficiency is connected to the monopoly’s high price: Consumers buy fewer units when the firm raises its price above marginal cost. But keep in mind that the profit earned on the units that continue to be sold is not the problem. The problem stems from the inefficiently low quantity of output. Put differently, if the high monopoly

---

**FIGURE 8**

*The Inefficiency of Monopoly*

Because a monopoly charges a price above marginal cost, not all consumers who value the good at more than its cost buy it. Thus, the quantity produced and sold by a monopoly is below the socially efficient level. The deadweight loss is represented by the area of the triangle between the demand curve (which reflects the value of the good to consumers) and the marginal-cost curve (which reflects the costs of the monopoly producer).
price did not discourage some consumers from buying the good, it would raise producer surplus by exactly the amount it reduced consumer surplus, leaving total surplus the same as could be achieved by a benevolent social planner.

There is, however, a possible exception to this conclusion. Suppose that a monopoly firm has to incur additional costs to maintain its monopoly position. For example, a firm with a government-created monopoly might need to hire lobbyists to convince lawmakers to continue its monopoly. In this case, the monopoly may use up some of its monopoly profits paying for these additional costs. If so, the social loss from monopoly includes both these costs and the deadweight loss resulting from a price above marginal cost.

**Quick Quiz** How does a monopolist’s quantity of output compare to the quantity of output that maximizes total surplus? How does this difference relate to the concept of deadweight loss?

### PRICE DISCRIMINATION

So far, we have been assuming that the monopoly firm charges the same price to all customers. Yet in many cases, firms sell the same good to different customers for different prices, even though the costs of producing for the two customers are the same. This practice is called **price discrimination**.

Before discussing the behavior of a price-discriminating monopolist, we should note that price discrimination is not possible when a good is sold in a competitive market. In a competitive market, many firms are selling the same good at the market price. No firm is willing to charge a lower price to any customer because the firm can sell all it wants at the market price. And if any firm tried to charge a higher price to a customer, that customer would buy from another firm. For a firm to price discriminate, it must have some market power.

### A PARABLE ABOUT PRICING

To understand why a monopolist would price discriminate, let’s consider an example. Imagine that you are the president of Readalot Publishing Company. Readalot’s best-selling author has just written a new novel. To keep things simple, let’s imagine that you pay the author a flat $2 million for the exclusive rights to publish the book. Let’s also assume that the cost of printing the book is zero. Readalot’s profit, therefore, is the revenue from selling the book minus the $2 million it has paid to the author. Given these assumptions, how would you, as Readalot’s president, decide the book’s price?

Your first step is to estimate the demand for the book. Readalot’s marketing department tells you that the book will attract two types of readers. The book will appeal to the author’s 100,000 die-hard fans who are willing to pay as much as $30. In addition, the book will appeal to about 400,000 less enthusiastic readers who will pay up to $5.

If Readalot charges a single price to all customers, what price maximizes profit? There are two natural prices to consider: $30 is the highest price Readalot can charge and still get the 100,000 die-hard fans, and $5 is the highest price it
can charge and still get the entire market of 500,000 potential readers. Solving Readalot’s problem is a matter of simple arithmetic. At a price of $30, Readalot sells 100,000 copies, has revenue of $3 million, and makes profit of $1 million. At a price of $5, it sells 500,000 copies, has revenue of $2.5 million, and makes profit of $500,000. Thus, Readalot maximizes profit by charging $30 and forgoing the opportunity to sell to the 400,000 less enthusiastic readers.

Notice that Readalot’s decision causes a deadweight loss. There are 400,000 readers willing to pay $5 for the book, and the marginal cost of providing it to them is zero. Thus, $2 million of total surplus is lost when Readalot charges the higher price. This deadweight loss is the inefficiency that arises whenever a monopolist charges a price above marginal cost.

Now suppose that Readalot’s marketing department makes a discovery: These two groups of readers are in separate markets. The die-hard fans live in Australia, and the other readers live in the United States. Moreover, it is hard for readers in one country to buy books in the other.

In response to this discovery, Readalot can change its marketing strategy and increase profits. To the 100,000 Australian readers, it can charge $30 for the book. To the 400,000 American readers, it can charge $5 for the book. In this case, revenue is $3 million in Australia and $2 million in the United States, for a total of $5 million. Profit is then $3 million, which is substantially greater than the $1 million the company could earn charging the same $30 price to all customers. Not surprisingly, Readalot chooses to follow this strategy of price discrimination.

The story of Readalot Publishing is hypothetical, but it describes accurately the business practice of many publishing companies. Textbooks, for example, are often sold at a lower price in Europe than in the United States. Even more important is the price differential between hardcover books and paperbacks. When a publisher has a new novel, it initially releases an expensive hardcover edition and later releases a cheaper paperback edition. The difference in price between these two editions far exceeds the difference in printing costs. The publisher’s goal is just as in our example. By selling the hardcover to die-hard fans and the paperback to less enthusiastic readers, the publisher price discriminates and raises its profit.

**The Moral of the Story**

Like any parable, the story of Readalot Publishing is stylized. Yet also like any parable, it teaches some general lessons. In this case, there are three lessons to be learned about price discrimination.

The first and most obvious lesson is that price discrimination is a rational strategy for a profit-maximizing monopolist. That is, by charging different prices to different customers, a monopolist can increase its profit. In essence, a price-discriminating monopolist charges each customer a price closer to his or her willingness to pay, therefore selling more than is possible with a single price.

The second lesson is that price discrimination requires the ability to separate customers according to their willingness to pay. In our example, customers were separated geographically. But sometimes monopolists choose other differences, such as age or income, to distinguish among customers.

A corollary to this second lesson is that certain market forces can prevent firms from price discriminating. In particular, one such force is arbitrage, the process of buying a good in one market at a low price and selling it in another market at
a higher price to profit from the price difference. In our example, if Australian bookstores could buy the book in the United States and resell it to Australian readers, the arbitrage would prevent Readalot from price discriminating, because no Australian would buy the book at the higher price.

The third lesson from our parable is the most surprising: Price discrimination can raise economic welfare. Recall that a deadweight loss arises when Readalot charges a single $30 price because the 400,000 less enthusiastic readers do not end up with the book, even though they value it at more than its marginal cost of production. By contrast, when Readalot price discriminates, all readers get the book, and the outcome is efficient. Thus, price discrimination can eliminate the inefficiency inherent in monopoly pricing.

Note that in this example the increase in welfare from price discrimination shows up as higher producer surplus rather than higher consumer surplus. Consumers are no better off for having bought the book: The price they pay exactly equals the value they place on the book, so they receive no consumer surplus. The entire increase in total surplus from price discrimination accrues to Readalot Publishing in the form of higher profit.

**The Analytics of Price Discrimination**

Let’s consider a bit more formally how price discrimination affects economic welfare. We begin by assuming that the monopolist can price discriminate perfectly. *Perfect price discrimination* describes a situation in which the monopolist knows exactly the willingness to pay of each customer and can charge each customer a different price. In this case, the monopolist charges each customer exactly his or her willingness to pay, and the monopolist gets the entire surplus in every transaction.

Figure 9 shows producer and consumer surplus with and without price discrimination. Without price discrimination, the firm charges a single price above marginal cost, as shown in panel (a). Because some potential customers who value the good at more than marginal cost do not buy it at this high price, the monopoly causes a deadweight loss. Yet when a firm can perfectly price discriminate, as shown in panel (b), each customer who values the good at more than marginal cost buys the good and is charged his or her willingness to pay. All mutually beneficial trades take place, there is no deadweight loss, and the entire surplus derived from the market goes to the monopoly producer in the form of profit.

In reality, of course, price discrimination is not perfect. Customers do not walk into stores with signs displaying their willingness to pay. Instead, firms price discriminate by dividing customers into groups: young versus old, weekday versus weekend shoppers, Americans versus Australians, and so on. Unlike those in our parable of Readalot Publishing, customers within each group differ in their willingness to pay, making perfect price discrimination impossible.

How does this imperfect price discrimination affect welfare? The analysis of these pricing schemes is quite complicated, and it turns out that there is no general answer to this question. Compared to the monopoly outcome with a single price, imperfect price discrimination can raise, lower, or leave unchanged total surplus in a market. The only certain conclusion is that price discrimination raises the monopoly’s profit; otherwise, the firm would choose to charge all customers the same price.
EXAMPLES OF PRICE DISCRIMINATION

Firms in our economy use various business strategies aimed at charging different prices to different customers. Now that we understand the economics of price discrimination, let’s consider some examples.

Movie Tickets  Many movie theaters charge a lower price for children and senior citizens than for other patrons. This fact is hard to explain in a competitive market. In a competitive market, price equals marginal cost, and the marginal cost of providing a seat for a child or senior citizen is the same as the marginal cost of providing a seat for anyone else. Yet the differential pricing is easily explained if movie theaters have some local monopoly power and if children and senior citizens have a lower willingness to pay for a ticket. In this case, movie theaters raise their profit by price discriminating.

Airline Prices  Seats on airplanes are sold at many different prices. Most airlines charge a lower price for a round-trip ticket between two cities if the traveler stays over a Saturday night. At first, this seems odd. Why should it matter to the airline whether a passenger stays over a Saturday night? The reason is that this rule provides a way to separate business travelers and leisure travelers. A passenger on a business trip has a high willingness to pay and, most likely, does not want to stay over a Saturday night. By contrast, a passenger traveling for personal reasons has

Panel (a) shows a monopolist that charges the same price to all customers. Total surplus in this market equals the sum of profit (producer surplus) and consumer surplus. Panel (b) shows a monopolist that can perfectly price discriminate. Because consumer surplus equals zero, total surplus now equals the firm’s profit. Comparing these two panels, you can see that perfect price discrimination raises profit, raises total surplus, and lowers consumer surplus.
a lower willingness to pay and is more likely to be willing to stay over a Saturday night. Thus, the airlines can successfully price discriminate by charging a lower price for passengers who stay over a Saturday night.

**Discount Coupons**  Many companies offer discount coupons to the public in newspapers and magazines. A buyer simply has to clip the coupon to get $0.50 off his or her next purchase. Why do companies offer these coupons? Why don’t they just cut the price of the product by $0.50?

The answer is that coupons allow companies to price discriminate. Companies know that not all customers are willing to spend the time to clip coupons. Moreover, the willingness to clip coupons is related to the customer’s willingness to pay for the good. A rich and busy executive is unlikely to spend her time clipping discount coupons out of the newspaper, and she is probably willing to pay a higher price for many goods. A person who is unemployed is more likely to clip coupons and to have a lower willingness to pay. Thus, by charging a lower price only to those customers who clip coupons, firms can successfully price discriminate.
Financial Aid  Many colleges and universities give financial aid to needy students. One can view this policy as a type of price discrimination. Wealthy students have greater financial resources and, therefore, a higher willingness to pay than needy students. By charging high tuition and selectively offering financial aid, schools in effect charge prices to customers based on the value they place on going to that school. This behavior is similar to that of any price-discriminating monopolist.

Quantity Discounts  So far in our examples of price discrimination, the monopolist charges different prices to different customers. Sometimes, however, monopolists price discriminate by charging different prices to the same customer for different units that the customer buys. For example, many firms offer lower prices to customers who buy large quantities. A bakery might charge $0.50 for each donut but $5 for a dozen. This is a form of price discrimination because the customer pays a higher price for the first unit bought than for the twelfth. Quantity discounts are often a successful way of price discriminating because a customer’s willingness to pay for an additional unit declines as the customer buys more units.
We have seen that monopolies, in contrast to competitive markets, fail to allocate resources efficiently. Monopolies produce less than the socially desirable quantity of output and, as a result, charge prices above marginal cost. Policymakers in the government can respond to the problem of monopoly in one of four ways:

- By trying to make monopolized industries more competitive
- By regulating the behavior of the monopolies
- By turning some private monopolies into public enterprises
- By doing nothing at all

**PUBLIC POLICY TOWARD MONOPOLIES**

Increasing Competition with Antitrust Laws

If Coca-Cola and PepsiCo wanted to merge, the deal would be closely examined by the federal government before it went into effect. The lawyers and economists in the Department of Justice might well decide that a merger between these two large soft drink companies would make the U.S. soft drink market substantially less competitive and, as a result, would reduce the economic well-being of the country as a whole. If so, the Department of Justice would challenge the merger in court, and if the judge agreed, the two companies would not be allowed to merge. It is precisely this kind of challenge that prevented software giant Microsoft from buying Intuit in 1994.

The government derives this power over private industry from the antitrust laws, a collection of statutes aimed at curbing monopoly power. The first and most important of these laws was the Sherman Antitrust Act, which Congress passed in 1890 to reduce the market power of the large and powerful “trusts” that were viewed as dominating the economy at the time. The Clayton Antitrust Act, passed in 1914, strengthened the government’s powers and authorized private lawsuits. As the U.S. Supreme Court once put it, the antitrust laws are “a comprehensive charter of economic liberty aimed at preserving free and unfettered competition as the rule of trade.”

The antitrust laws give the government various ways to promote competition. They allow the government to prevent mergers, such as our hypothetical merger between Coca-Cola and PepsiCo. They also allow the government to break up companies. For example, in 1984, the government split up AT&T, the large telecommunications company, into eight smaller companies. Finally, the antitrust laws prevent companies from coordinating their activities in ways that make markets less competitive.

Antitrust laws have costs as well as benefits. Sometimes companies merge not to reduce competition but to lower costs through more efficient joint production. These benefits from mergers are sometimes called *synergies*. For example, many U.S. banks have merged in recent years and, by combining operations, have been able to reduce administrative staff. If antitrust laws are to raise social welfare, the government must be able to determine which mergers are desirable and which are not. That is, it must be able to measure and compare the social benefit from
synergies to the social costs of reduced competition. Critics of the antitrust laws are skeptical that the government can perform the necessary cost–benefit analysis with sufficient accuracy.

**Regulation**

Another way the government deals with the problem of monopoly is by regulating the behavior of monopolists. This solution is common in the case of natural monopolies, such as water and electric companies. These companies are not allowed to charge any price they want. Instead, government agencies regulate their prices.

What price should the government set for a natural monopoly? This question is not as easy as it might at first appear. One might conclude that the price should equal the monopolist’s marginal cost. If price equals marginal cost, customers will buy the quantity of the monopolist’s output that maximizes total surplus, and the allocation of resources will be efficient.
There are, however, two practical problems with marginal-cost pricing as a regulatory system. The first arises from the logic of cost curves. By definition, natural monopolies have declining average total cost. As we first discussed in Chapter 12, when average total cost is declining, marginal cost is less than average total cost. This situation is illustrated in Figure 10, which shows a firm with a large fixed cost and then constant marginal cost thereafter. If regulators were to set price equal to marginal cost, that price must be less than the firm’s average total cost, and the firm would lose money. Instead of charging such a low price, the monopoly firm would just exit the industry.

Regulators can respond to this problem in various ways, none of which is perfect. One way is to subsidize the monopolist. In essence, the government picks
up the losses inherent in marginal-cost pricing. Yet to pay for the subsidy, the
government needs to raise money through taxation, which involves its own dead-
weight losses. Alternatively, the regulators can allow the monopolist to charge a
price higher than marginal cost. If the regulated price equals average total cost,
the monopolist earns exactly zero economic profit. Yet average-cost pricing leads
to deadweight losses because the monopolist’s price no longer reflects the mar-
ginal cost of producing the good. In essence, average-cost pricing is like a tax on
the good the monopolist is selling.

The second problem with marginal-cost pricing as a regulatory system (and
with average-cost pricing as well) is that it gives the monopolist no incentive to
reduce costs. Each firm in a competitive market tries to reduce its costs because
lower costs mean higher profits. But if a regulated monopolist knows that regulators will reduce prices whenever costs fall, the monopolist will not benefit from lower costs. In practice, regulators deal with this problem by allowing monopolists to keep some of the benefits from lower costs in the form of higher profit, a practice that requires some departure from marginal-cost pricing.

**Public Ownership**

The third policy used by the government to deal with monopoly is public ownership. That is, rather than regulating a natural monopoly that is run by a private firm, the government can run the monopoly itself. This solution is common in many European countries, where the government owns and operates utilities such as telephone, water, and electric companies. In the United States, the government runs the Postal Service. The delivery of ordinary first-class mail is often thought to be a natural monopoly.

Economists usually prefer private to public ownership of natural monopolies. The key issue is how the ownership of the firm affects the costs of production. Private owners have an incentive to minimize costs as long as they reap part of the benefit in the form of higher profit. If the firm’s managers are doing a bad job of keeping costs down, the firm’s owners will fire them. By contrast, if the government bureaucrats who run a monopoly do a bad job, the losers are the customers and taxpayers, whose only recourse is the political system. The bureaucrats may become a special-interest group and attempt to block cost-reducing reforms. Put simply, as a way of ensuring that firms are well run, the voting booth is less reliable than the profit motive.

**Doing Nothing**

Each of the foregoing policies aimed at reducing the problem of monopoly has drawbacks. As a result, some economists argue that it is often best for the gov-
A famous theorem in economics states that a competitive enterprise economy will produce the largest possible income from a given stock of resources. No real economy meets the exact conditions of the theorem, and all real economies will fall short of the ideal economy—a difference called “market failure.” In my view, however, the degree of “market failure” for the American economy is much smaller than the “political failure” arising from the imperfections of economic policies found in real political systems.

As this quotation makes clear, determining the proper role of the government in the economy requires judgments about politics as well as economics.

QUICK QUIZ Describe the ways policymakers can respond to the inefficiencies caused by monopolies. List a potential problem with each of these policy responses.

**CONCLUSION: THE PREVALENCE OF MONOPOLIES**

This chapter has discussed the behavior of firms that have control over the prices they charge. We have seen that these firms behave very differently from the competitive firms studied in the previous chapter. Table 2 summarizes some of the key similarities and differences between competitive and monopoly markets.

From the standpoint of public policy, a crucial result is that a monopolist produces less than the socially efficient quantity and charges a price above marginal cost. As a result, a monopoly causes deadweight losses. In some cases, these

---

**TABLE 2**

<table>
<thead>
<tr>
<th>Similarities</th>
<th>Competition</th>
<th>Monopoly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal of firms</td>
<td>Maximize profits</td>
<td>Maximize profits</td>
</tr>
<tr>
<td>Rule for maximizing</td>
<td>$MR = MC$</td>
<td>$MR = MC$</td>
</tr>
<tr>
<td>Can earn economic profits in the short run?</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Differences</th>
<th>Competition</th>
<th>Monopoly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of firms</td>
<td>Many</td>
<td>One</td>
</tr>
<tr>
<td>Marginal revenue</td>
<td>$MR = P$</td>
<td>$MR &lt; P$</td>
</tr>
<tr>
<td>Price</td>
<td>$P = MC$</td>
<td>$P &gt; MC$</td>
</tr>
<tr>
<td>Produces welfare-maximizing level of output?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Entry in long run?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Can earn economic profits in long run?</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Price discrimination possible?</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Competition versus Monopoly: A Summary Comparison**
inefficiencies can be mitigated through price discrimination by the monopolist, but other times, they call for policymakers to take an active role.

How prevalent are the problems of monopoly? There are two answers to this question.

In one sense, monopolies are common. Most firms have some control over the prices they charge. They are not forced to charge the market price for their goods because their goods are not exactly the same as those offered by other firms. A Ford Taurus is not the same as a Toyota Camry. Ben and Jerry’s ice cream is not the same as Breyer’s. Each of these goods has a downward-sloping demand curve, which gives each producer some degree of monopoly power.

Yet firms with substantial monopoly power are rare. Few goods are truly unique. Most have substitutes that, even if not exactly the same, are similar. Ben and Jerry can raise the price of their ice cream a little without losing all their sales, but if they raise it very much, sales will fall substantially as their customers switch to another brand.

In the end, monopoly power is a matter of degree. It is true that many firms have some monopoly power. It is also true that their monopoly power is usually limited. In such a situation, we will not go far wrong assuming that firms operate in competitive markets, even if that is not precisely the case.

**SUMMARY**

- A monopoly is a firm that is the sole seller in its market. A monopoly arises when a single firm owns a key resource, when the government gives a firm the exclusive right to produce a good, or when a single firm can supply the entire market at a smaller cost than many firms could.
- Because a monopoly is the sole producer in its market, it faces a downward-sloping demand curve for its product. When a monopoly increases production by 1 unit, it causes the price of its good to fall, which reduces the amount of revenue earned on all units produced. As a result, a monopoly’s marginal revenue is always below the price of its good.
- Like a competitive firm, a monopoly firm maximizes profit by producing the quantity at which marginal revenue equals marginal cost. The monopoly then chooses the price at which that quantity is demanded. Unlike a competitive firm, a monopoly firm’s price exceeds its marginal revenue, so its price exceeds marginal cost.
- A monopolist’s profit-maximizing level of output is below the level that maximizes the sum of consumer and producer surplus. That is, when the monopoly charges a price above marginal cost, some consumers who value the good more than its cost of production do not buy it. As a result, monopoly causes deadweight losses similar to the deadweight losses caused by taxes.
- A monopolist often can raise its profits by charging different prices for the same good based on a buyer’s willingness to pay. This practice of price discrimination can raise economic welfare by getting the good to some consumers who otherwise would not buy it. In the extreme case of perfect price discrimination, the deadweight loss of monopoly is completely eliminated, and all the surplus in the market goes to the monopoly producer. More generally, when price discrimination is imperfect, it can either raise or lower welfare compared to the outcome with a single monopoly price.
• Policymakers can respond to the inefficiency of monopoly behavior in four ways. They can use the antitrust laws to try to make the industry more competitive. They can regulate the prices that the monopoly charges. They can turn the monopolist into a government-run enterprise. Or if the market failure is deemed small compared to the inevitable imperfections of policies, they can do nothing at all.

**KEY CONCEPTS**

monopoly, p. 288  
natural monopoly, p. 290  
price discrimination, p. 302

**QUESTIONS FOR REVIEW**

1. Give an example of a government-created monopoly. Is creating this monopoly necessarily bad public policy? Explain.
2. Define natural monopoly. What does the size of a market have to do with whether an industry is a natural monopoly?
3. Why is a monopolist’s marginal revenue less than the price of its good? Can marginal revenue ever be negative? Explain.
4. Draw the demand, marginal-revenue, average-total-cost, and marginal-cost curves for a monopolist. Show the profit-maximizing level of output, the profit-maximizing price, and the amount of profit.
5. In your diagram from the previous question, show the level of output that maximizes total surplus. Show the deadweight loss from the monopoly. Explain your answer.
6. Give two examples of price discrimination. In each case, explain why the monopolist chooses to follow this business strategy.
7. What gives the government the power to regulate mergers between firms? From the standpoint of the welfare of society, give a good reason and a bad reason that two firms might want to merge.
8. Describe the two problems that arise when regulators tell a natural monopoly that it must set a price equal to marginal cost.
1. A publisher faces the following demand schedule for the next novel of one of its popular authors:

<table>
<thead>
<tr>
<th>Price</th>
<th>Quantity Demanded</th>
</tr>
</thead>
<tbody>
<tr>
<td>$100</td>
<td>0 novels</td>
</tr>
<tr>
<td>90</td>
<td>100,000</td>
</tr>
<tr>
<td>80</td>
<td>200,000</td>
</tr>
<tr>
<td>70</td>
<td>300,000</td>
</tr>
<tr>
<td>60</td>
<td>400,000</td>
</tr>
<tr>
<td>50</td>
<td>500,000</td>
</tr>
<tr>
<td>40</td>
<td>600,000</td>
</tr>
<tr>
<td>30</td>
<td>700,000</td>
</tr>
<tr>
<td>20</td>
<td>800,000</td>
</tr>
<tr>
<td>10</td>
<td>900,000</td>
</tr>
<tr>
<td>0</td>
<td>1,000,000</td>
</tr>
</tbody>
</table>

The author is paid $2 million to write the book, and the marginal cost of publishing the book is a constant $10 per book.

a. Compute total revenue, total cost, and profit at each quantity. What quantity would a profit-maximizing publisher choose? What price would it charge?
b. Compute marginal revenue. (Recall that \( MR = \frac{\Delta TR}{\Delta Q} \).) How does marginal revenue compare to the price? Explain.
c. Graph the marginal-revenue, marginal-cost, and demand curves. At what quantity do the marginal-revenue and marginal-cost curves cross? What does this signify?
d. In your graph, shade in the deadweight loss. Explain in words what this means.
e. If the author were paid $3 million instead of $2 million to write the book, how would this affect the publisher’s decision regarding the price to charge? Explain.
f. Suppose the publisher was not profit-maximizing but was concerned with maximizing economic efficiency. What price would it charge for the book? How much profit would it make at this price?

2. Suppose that a natural monopolist was required by law to charge average total cost. On a diagram, label the price charged and the deadweight loss to society relative to marginal-cost pricing.

3. Suppose the Clean Springs Water Company has a monopoly on bottled water sales in California. If the price of tap water increases, what is the change in Clean Springs’ profit-maximizing levels of output, price, and profit? Explain in words and with a graph.

4. A small town is served by many competing supermarkets, which have constant marginal cost.

a. Using a diagram of the market for groceries, show the consumer surplus, producer surplus, and total surplus.
b. Now suppose that the independent supermarkets combine into one chain. Using a new diagram, show the new consumer surplus, producer surplus, and total surplus. Relative to the competitive market, what is the transfer from consumers to producers? What is the deadweight loss?

5. Johnny Rockabilly has just finished recording his latest CD. His record company’s marketing department determines that the demand for the CD is as follows:

<table>
<thead>
<tr>
<th>Price</th>
<th>Number of CDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$24</td>
<td>10,000</td>
</tr>
<tr>
<td>22</td>
<td>20,000</td>
</tr>
<tr>
<td>20</td>
<td>30,000</td>
</tr>
<tr>
<td>18</td>
<td>40,000</td>
</tr>
<tr>
<td>16</td>
<td>50,000</td>
</tr>
<tr>
<td>14</td>
<td>60,000</td>
</tr>
</tbody>
</table>

The company can produce the CD with no fixed cost and a variable cost of $5 per CD.
a. Find total revenue for quantity equal to 10,000, 20,000, and so on. What is the marginal revenue for each 10,000 increase in the quantity sold?

b. What quantity of CDs would maximize profit? What would the price be? What would the profit be?

c. If you were Johnny’s agent, what recording fee would you advise Johnny to demand from the record company? Why?

6. A company is considering building a bridge across a river. The bridge would cost $2 million to build and nothing to maintain. The following table shows the company’s anticipated demand over the lifetime of the bridge:

<table>
<thead>
<tr>
<th>Price per Crossing</th>
<th>Number of Crossings, in Thousands</th>
</tr>
</thead>
<tbody>
<tr>
<td>$8</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>200</td>
</tr>
<tr>
<td>5</td>
<td>300</td>
</tr>
<tr>
<td>4</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>500</td>
</tr>
<tr>
<td>2</td>
<td>600</td>
</tr>
<tr>
<td>1</td>
<td>700</td>
</tr>
<tr>
<td>0</td>
<td>800</td>
</tr>
</tbody>
</table>

a. If the company were to build the bridge, what would be its profit-maximizing price? Would that be the efficient level of output? Why or why not?

b. If the company is interested in maximizing profit, should it build the bridge? What would be its profit or loss?

c. If the government were to build the bridge, what price should it charge?

d. Should the government build the bridge? Explain.

7. Larry, Curly, and Moe run the only saloon in town. Larry wants to sell as many drinks as possible without losing money. Curly wants the saloon to bring in as much revenue as possible. Moe wants to make the largest possible profits. Using a single diagram of the saloon’s demand curve and its cost curves, show the price and quantity combinations favored by each of the three partners. Explain.

8. For many years, AT&T was a regulated monopoly, providing both local and long-distance telephone service.

a. Explain why long-distance phone service was originally a natural monopoly.

b. Over the past two decades, many companies have launched communication satellites, each of which can transmit a limited number of calls. How did the growing role of satellites change the cost structure of long-distance phone service?

After a lengthy legal battle with the government, AT&T agreed to compete with other companies in the long-distance market. It also agreed to spin off its local phone service into the “Baby Bells,” which remain highly regulated.

c. Why might it be efficient to have competition in long-distance phone service and regulated monopolies in local phone service?

9. Consider the relationship between monopoly pricing and price elasticity of demand:

a. Explain why a monopolist will never produce a quantity at which the demand curve is inelastic. (Hint: If demand is inelastic and the firm raises its price, what happens to total revenue and total costs?)

b. Draw a diagram for a monopolist, precisely labeling the portion of the demand curve that is inelastic. (Hint: The answer is related to the marginal-revenue curve.)

c. On your diagram, show the quantity and price that maximizes total revenue.

10. If the government wanted to encourage a monopoly to produce the socially efficient quantity, should it use a per-unit tax or a per-unit subsidy? Explain how this tax or subsidy would achieve the socially efficient level of output. Among the various interested parties—the monopoly firm, the monopoly’s consumers, and other taxpayers—who would support the policy and who would oppose it?
11. You live in a town with 300 adults and 200 children, and you are thinking about putting on a play to entertain your neighbors and make some money. A play has a fixed cost of $2,000, but selling an extra ticket has zero marginal cost. Here are the demand schedules for your two types of customer:

<table>
<thead>
<tr>
<th>Price</th>
<th>Adults</th>
<th>Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>300</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>300</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>300</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>300</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>300</td>
<td>200</td>
</tr>
<tr>
<td>1</td>
<td>300</td>
<td>200</td>
</tr>
<tr>
<td>0</td>
<td>300</td>
<td>200</td>
</tr>
</tbody>
</table>

a. To maximize profit, what price would you charge for an adult ticket? For a child’s ticket? How much profit do you make?
b. The city council passes a law prohibiting you from charging different prices to different customers. What price do you set for a ticket now? How much profit do you make?
c. Who is worse off because of the law prohibiting price discrimination? Who is better off? (If you can, quantify the changes in welfare.)
d. If the fixed cost of the play were $2,500 rather than $2,000, how would your answers to parts (a), (b), and (c) change?

12. Based on market research, a recording company obtains the following information about the demand and production costs of its new CD:

\[
\begin{align*}
\text{Price} &= 1,000 - 10Q \\
\text{Total Revenue} &= 1,000Q - 10Q^2 \\
\text{Marginal Revenue} &= 1,000 - 20Q \\
\text{Marginal Cost} &= 100 + 10Q
\end{align*}
\]

where \(Q\) indicates the number of copies sold and \(P\) is the price in cents.

a. Find the price and quantity that maximizes the company’s profit.
b. Find the price and quantity that would maximize social welfare.
c. Calculate the deadweight loss from monopoly.
d. Suppose, in addition to the costs above, the musician on the album has to be paid. The company is considering four options:
i. A flat fee of 2,000 cents
ii. 50 percent of the profits
iii. 150 cents per unit sold
iv. 50 percent of the revenue
For each option, calculate the profit-maximizing price and quantity. Which, if any, of these compensation schemes would alter the deadweight loss from monopoly? Explain.

13. Many schemes for price discriminating involve some cost. For example, discount coupons take up the time and resources of both the buyer and the seller. This question considers the implications of costly price discrimination. To keep things simple, let’s assume that our monopolist’s production costs are simply proportional to output so that average total cost and marginal cost are constant and equal to each other.

a. Draw the cost, demand, and marginal-revenue curves for the monopolist. Show the price the monopolist would charge without price discrimination.
b. In your diagram, mark the area equal to the monopolist’s profit and call it \(X\). Mark the area equal to consumer surplus and call it \(Y\). Mark the area equal to the deadweight loss and call it \(Z\).
c. Now suppose that the monopolist can perfectly price discriminate. What is the monopolist’s profit? (Give your answer in terms of \(X, Y,\) and \(Z\).)
d. What is the change in the monopolist’s profit from price discrimination? What is the change in total surplus from price discrimination? Which change is larger? Explain. (Give your answer in terms of \(X, Y,\) and \(Z\).)
e. Now suppose that there is some cost of price discrimination. To model this cost, let’s assume that the monopolist has to pay a fixed cost $C$ to price discriminate. How would a monopolist make the decision whether to pay this fixed cost? (Give your answer in terms of $X$, $Y$, $Z$, and $C$.)

f. How would a benevolent social planner, who cares about total surplus, decide whether the monopolist should price discriminate? (Give your answer in terms of $X$, $Y$, $Z$, and $C$.)

g. Compare your answers to parts (e) and (f). How does the monopolist’s incentive to price discriminate differ from the social planner’s? Is it possible that the monopolist will price discriminate even though it is not socially desirable?
PART VI

The Data of Macroeconomics
Measuring a Nation’s Income

When you finish school and start looking for a full-time job, your experience will, to a large extent, be shaped by prevailing economic conditions. In some years, firms throughout the economy are expanding their production of goods and services, employment is rising, and jobs are easy to find. In other years, firms are cutting back production, employment is declining, and finding a good job takes a long time. Not surprisingly, any college graduate would rather enter the labor force in a year of economic expansion than in a year of economic contraction.

Because the health of the overall economy profoundly affects all of us, changes in economic conditions are widely reported by the media. Indeed, it is hard to pick up a newspaper, check an online news service, or turn on the TV without seeing some newly reported statistic about the economy. The statistic might measure the total income of everyone in the economy (GDP), the rate at which average prices are rising (inflation), the percentage of the labor force that is out of work (unemployment), total spending at stores (retail sales), or the imbalance of trade between the United States and the rest of the world (the trade deficit). All these statistics are macroeconomic. Rather than telling us about a particular household, firm, or market, they tell us something about the entire economy.

As you may recall from Chapter 2, economics is divided into two branches: microeconomics and macroeconomics. Microeconomics is the study of how individual households and firms make decisions and how they interact with one
Macroeconomics is the study of the economy as a whole. The goal of macroeconomics is to explain the economic changes that affect many households, firms, and markets simultaneously. Macroeconomists address diverse questions: Why is average income high in some countries while it is low in others? Why do prices sometimes rise rapidly while at other times they are more stable? Why do production and employment expand in some years and contract in others? What, if anything, can the government do to promote rapid growth in incomes, low inflation, and stable employment? These questions are all macroeconomic in nature because they concern the workings of the entire economy.

Because the economy as a whole is just a collection of many households and many firms interacting in many markets, microeconomics and macroeconomics are closely linked. The basic tools of supply and demand, for instance, are as central to macroeconomic analysis as they are to microeconomic analysis. Yet studying the economy in its entirety raises some new and intriguing challenges.

In this and the next chapter, we discuss some of the data that economists and policymakers use to monitor the performance of the overall economy. These data reflect the economic changes that macroeconomists try to explain. This chapter considers gross domestic product, or simply GDP, which measures the total income of a nation. GDP is the most closely watched economic statistic because it is thought to be the best single measure of a society’s economic well-being.

**THE ECONOMY’S INCOME AND EXPENDITURE**

If you were to judge how a person is doing economically, you might first look at his or her income. A person with a high income can more easily afford life’s necessities and luxuries. It is no surprise that people with higher incomes enjoy higher standards of living—better housing, better healthcare, fancier cars, more opulent vacations, and so on.

The same logic applies to a nation’s overall economy. When judging whether the economy is doing well or poorly, it is natural to look at the total income that everyone in the economy is earning. That is the task of gross domestic product (GDP).

GDP measures two things at once: the total income of everyone in the economy and the total expenditure on the economy’s output of goods and services. GDP can perform the trick of measuring both total income and total expenditure because these two things are really the same. For an economy as a whole, income must equal expenditure.

Why is this true? An economy’s income is the same as its expenditure because every transaction has two parties: a buyer and a seller. Every dollar of spending by some buyer is a dollar of income for some seller. Suppose, for instance, that Karen pays Doug $100 to mow her lawn. In this case, Doug is a seller of a service, and Karen is a buyer. Doug earns $100, and Karen spends $100. Thus, the transaction contributes equally to the economy’s income and to its expenditure. GDP, whether measured as total income or total expenditure, rises by $100.

Another way to see the equality of income and expenditure is with the circular-flow diagram in Figure 1. As you may recall from Chapter 2, this diagram describes all the transactions between households and firms in a simple economy. It simplifies matters by assuming that all goods and services are bought by households
and that households spend all of their income. In this economy, when households buy goods and services from firms, these expenditures flow through the markets for goods and services. When the firms in turn use the money they receive from sales to pay workers’ wages, landowners’ rent, and firm owners’ profit, this income flows through the markets for the factors of production. Money continuously flows from households to firms and then back to households.

GDP measures this flow of money. We can compute it for this economy in one of two ways: by adding up the total expenditure by households or by adding up the total income (wages, rent, and profit) paid by firms. Because all expenditure in the economy ends up as someone’s income, GDP is the same regardless of how we compute it.

The actual economy is, of course, more complicated than the one illustrated in Figure 1. Households do not spend all of their income; they pay some of it to the government in taxes, and they save some for use in the future. In addition, households do not buy all goods and services produced in the economy; some goods and services are bought by governments, and some are bought by firms that plan to use them in the future to produce their own output. Yet the basic lesson remains the same: Regardless of whether a household, government, or firm buys a good or service, the transaction has a buyer and seller. Thus, for the economy as a whole, expenditure and income are always the same.
Having discussed the meaning of gross domestic product in general terms, let’s be more precise about how this statistic is measured. Here is a definition of GDP that focuses on GDP as a measure of total expenditure:

- **Gross domestic product (GDP)** is the market value of all final goods and services produced within a country in a given period of time.

This definition might seem simple enough. But in fact, many subtle issues arise when computing an economy’s GDP. Let’s therefore consider each phrase in this definition with some care.

**“GDP IS THE MARKET VALUE . . .”**

You have probably heard the adage, “You can’t compare apples and oranges.” Yet GDP does exactly that. GDP adds together many different kinds of products into a single measure of the value of economic activity. To do this, it uses market prices. Because market prices measure the amount people are willing to pay for different goods, they reflect the value of those goods. If the price of an apple is twice the price of an orange, then an apple contributes twice as much to GDP as does an orange.

**“. . . OF ALL . . .”**

GDP tries to be comprehensive. It includes all items produced in the economy and sold legally in markets. GDP measures the market value of not just apples and oranges but also pears and grapefruit, books and movies, haircuts and healthcare, and on and on.

GDP also includes the market value of the housing services provided by the economy’s stock of housing. For rental housing, this value is easy to calculate—the rent equals both the tenant’s expenditure and the landlord’s income. Yet many people own the place where they live and, therefore, do not pay rent. The government includes this owner-occupied housing in GDP by estimating its rental value. In effect, GDP is based on the assumption that the owner is renting the house to himself. The imputed rent is included both in the homeowner’s expenditure and in his income, so it adds to GDP.

There are some products, however, that GDP excludes because measuring them is so difficult. GDP excludes most items produced and sold illicitly, such as illegal drugs. It also excludes most items that are produced and consumed at home and, therefore, never enter the marketplace. Vegetables you buy at the grocery store are part of GDP; vegetables you grow in your garden are not.

These exclusions from GDP can at times lead to paradoxical results. For example, when Karen pays Doug to mow her lawn, that transaction is part of GDP. If Karen were to marry Doug, the situation would change. Even though Doug may continue to mow Karen’s lawn, the value of the mowing is now left out of GDP because Doug’s service is no longer sold in a market. Thus, when Karen and Doug marry, GDP falls.
“... **Final ...**”

When International Paper makes paper, which Hallmark then uses to make a greeting card, the paper is called an *intermediate good*, and the card is called a *final good*. GDP includes only the value of final goods. This is done because the value of intermediate goods is already included in the prices of the final goods. Adding the market value of the paper to the market value of the card would be double counting. That is, it would (incorrectly) count the paper twice.

An important exception to this principle arises when an intermediate good is produced and, rather than being used, is added to a firm’s inventory of goods for use or sale at a later date. In this case, the intermediate good is taken to be “final” for the moment, and its value as inventory investment is included as part of GDP. Thus, additions to inventory add to GDP, and when the goods in inventory are later used or sold, the reductions in inventory subtract from GDP.

“... **Goods and Services ...**”

GDP includes both tangible goods (food, clothing, cars) and intangible services (haircuts, housecleaning, doctor visits). When you buy a CD by your favorite band, you are buying a good, and the purchase price is part of GDP. When you pay to hear a concert by the same band, you are buying a service, and the ticket price is also part of GDP.

“... **Produced ...**”

GDP includes goods and services currently produced. It does not include transactions involving items produced in the past. When General Motors produces and sells a new car, the value of the car is included in GDP. When one person sells a used car to another person, the value of the used car is not included in GDP.

“... **Within a Country ...**”

GDP measures the value of production within the geographic confines of a country. When a Canadian citizen works temporarily in the United States, her production is part of U.S. GDP. When an American citizen owns a factory in Haiti, the production at his factory is not part of U.S. GDP. (It is part of Haiti’s GDP.) Thus, items are included in a nation’s GDP if they are produced domestically, regardless of the nationality of the producer.

“... **In a Given Period of Time.**”

GDP measures the value of production that takes place within a specific interval of time. Usually, that interval is a year or a quarter (three months). GDP measures the economy’s flow of income and expenditure during that interval.

When the government reports the GDP for a quarter, it usually presents GDP “at an annual rate.” This means that the figure reported for quarterly GDP is the amount of income and expenditure during the quarter multiplied by 4. The government uses this convention so that quarterly and annual figures on GDP can be compared more easily.

In addition, when the government reports quarterly GDP, it presents the data after they have been modified by a statistical procedure called *seasonal adjustment*. 
The unadjusted data show clearly that the economy produces more goods and services during some times of year than during others. (As you might guess, December’s holiday shopping season is a high point.) When monitoring the condition of the economy, economists and policymakers often want to look beyond these regular seasonal changes. Therefore, government statisticians adjust the quarterly data to take out the seasonal cycle. The GDP data reported in the news are always seasonally adjusted.

Now let’s repeat the definition of GDP:

- Gross domestic product (GDP) is the market value of all final goods and services produced within a country in a given period of time.

This definition focuses on GDP as total expenditure in the economy. But don’t forget that every dollar spent by a buyer of a good or service becomes a dollar of income to the seller of that good or service. Therefore, in addition to applying this definition, the government adds up total income in the economy. The two ways of calculating GDP give almost exactly the same answer. (Why “almost”? Although the two measures should be precisely the same, data sources are not perfect. The difference between the two calculations of GDP is called the statistical discrepancy.)

It should be apparent that GDP is a sophisticated measure of the value of economic activity. In advanced courses in macroeconomics, you will learn more about the subtleties that arise in its calculation. But even now you can see that each phrase in this definition is packed with meaning.

**Quick Quiz** Which contributes more to GDP—the production of a pound of hamburger or the production of a pound of caviar? Why?

---

**THE COMPONENTS OF GDP**

Spending in the economy takes many forms. At any moment, the Smith family may be having lunch at Burger King; General Motors may be building a car factory; the Navy may be procuring a submarine; and British Airways may be buying an airplane from Boeing. GDP includes all of these various forms of spending on domestically produced goods and services.

To understand how the economy is using its scarce resources, economists study the composition of GDP among various types of spending. To do this, GDP (which we denote as \( \text{\$} \)) is divided into four components: consumption (\( C \)), investment (\( I \)), government purchases (\( G \)), and net exports (\( NX \)):

\[
Y = C + I + G + NX.
\]

This equation is an identity—an equation that must be true because of how the variables in the equation are defined. In this case, because each dollar of expenditure included in GDP is placed into one of the four components of GDP, the total of the four components must be equal to GDP. Let’s look at each of these four components more closely.
CONSUMPTION

Consumption is spending by households on goods and services, with the exception of purchases of new housing. Goods include household spending on durable goods, such as automobiles and appliances, and nondurable goods, such as food and clothing. Services include such intangible items as haircuts and medical care. Household spending on education is also included in consumption of services (although one might argue that it would fit better in the next component).

INVESTMENT

Investment is the purchase of goods that will be used in the future to produce more goods and services. It is the sum of purchases of capital equipment, inventories, and structures, including household purchases of new housing.
and structures. Investment in structures includes expenditure on new housing. By convention, the purchase of a new house is the one form of household spending categorized as investment rather than consumption.

As mentioned earlier in this chapter, the treatment of inventory accumulation is noteworthy. When Dell produces a computer and adds it to its inventory instead of selling it, Dell is assumed to have “purchased” the computer for itself. That is, the national income accountants treat the computer as part of Dell’s investment spending. (If Dell later sells the computer out of inventory, Dell’s inventory investment will then be negative, offsetting the positive expenditure of the buyer.) Inventories are treated this way because one aim of GDP is to measure the value of the economy’s production, and goods added to inventory are part of that period’s production.

Notice that GDP accounting uses the word investment differently from how you might hear the term in everyday conversation. When you hear the word investment, you might think of financial investments, such as stocks, bonds, and mutual funds—topics that we study later in this book. By contrast, because GDP measures expenditure on goods and services, here the word investment means purchases of goods (such as capital equipment, structures, and inventories) used to produce other goods.

**Government Purchases**

Government purchases include spending on goods and services by local, state, and federal governments. It includes the salaries of government workers as well as expenditures on public works. Recently, the U.S. national income accounts have switched to the longer label government consumption expenditure and gross investment, but in this book, we will use the traditional and shorter term government purchases.

The meaning of government purchases requires a bit of clarification. When the government pays the salary of an Army general or a schoolteacher, that salary is part of government purchases. But when the government pays a Social Security benefit to a person who is elderly or an unemployment insurance benefit to a worker who was recently laid off, the story is very different: These are called transfer payments because they are not made in exchange for a currently produced good or service. Transfer payments alter household income, but they do not reflect the economy’s production. (From a macroeconomic standpoint, transfer payments are like negative taxes.) Because GDP is intended to measure income from, and expenditure on, the production of goods and services, transfer payments are not counted as part of government purchases.

**Net Exports**

Net exports equal the foreign purchases of domestically produced goods (exports) minus the domestic purchases of foreign goods (imports). A domestic firm’s sale to a buyer in another country, such as Boeing’s sale of an airplane to British Airways, increases net exports.

The net in net exports refers to the fact that imports are subtracted from exports. This subtraction is made because other components of GDP include imports of goods and services. For example, suppose that a household buys a $30,000 car from Volvo, the Swedish carmaker. That transaction increases consumption by $30,000 because car purchases are part of consumer spending. It also reduces...
net exports by $30,000 because the car is an import. In other words, net exports include goods and services produced abroad (with a minus sign) because these goods and services are included in consumption, investment, and government purchases (with a plus sign). Thus, when a domestic household, firm, or government buys a good or service from abroad, the purchase reduces net exports, but because it also raises consumption, investment, or government purchases, it does not affect GDP.

THE COMPONENTS OF U.S. GDP

Table 1 shows the composition of U.S. GDP in 2007. In this year, the GDP of the United States was almost $14 trillion. Dividing this number by the 2007 U.S. population of 302 million yields GDP per person (sometimes called GDP per capita). We find that in 2007 the income and expenditure of the average American was $45,838.

Consumption made up 70 percent of GDP, or $32,225 per person. Investment was $7,061 per person. Government purchases were $8,912 per person. Net exports were –$2,360 per person. This number is negative because Americans earned less from selling to foreigners than they spent on foreign goods.

These data come from the Bureau of Economic Analysis, which is the part of the U.S. Department of Commerce that produces the national income accounts. You can find more recent data on GDP at its website, http://www.bea.gov.

QUICK QUIZ
List the four components of expenditure. Which is the largest?

REAL VERSUS NOMINAL GDP

As we have seen, GDP measures the total spending on goods and services in all markets in the economy. If total spending rises from one year to the next, at least one of two things must be true: (1) the economy is producing a larger output of goods and services, or (2) goods and services are being sold at higher prices.

When studying changes in the economy over time, economists want to separate these two effects. In particular, they want a measure of the total quantity of goods
and services the economy is producing that is not affected by changes in the prices of those goods and services.

To do this, economists use a measure called real GDP. Real GDP answers a hypothetical question: What would be the value of the goods and services produced this year if we valued these goods and services at the prices that prevailed in some specific year in the past? By evaluating current production using prices that are fixed at past levels, real GDP shows how the economy’s overall production of goods and services changes over time.

To see more precisely how real GDP is constructed, let’s consider an example.

**A Numerical Example**

Table 2 shows some data for an economy that produces only two goods: hot dogs and hamburgers. The table shows the prices and quantities produced of the two goods in the years 2008, 2009, and 2010.

To compute total spending in this economy, we would multiply the quantities of hot dogs and hamburgers by their prices. In the year 2008, 100 hot dogs are sold at a price of $1 per hot dog, so expenditure on hot dogs equals $100. In the same year, 50 hamburgers are sold for $2 per hamburger, so expenditure on hamburgers also equals $100. Total expenditure in the economy—the sum of expenditure on hot dogs and expenditure on hamburgers—is $200. This amount, the production of goods and services valued at current prices, is called nominal GDP.

<table>
<thead>
<tr>
<th>Year</th>
<th>Price of Hot Dogs</th>
<th>Quantity of Hot Dogs</th>
<th>Price of Hamburgers</th>
<th>Quantity of Hamburgers</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>$1</td>
<td>100</td>
<td>$2</td>
<td>50</td>
</tr>
<tr>
<td>2009</td>
<td>$2</td>
<td>150</td>
<td>$3</td>
<td>100</td>
</tr>
<tr>
<td>2010</td>
<td>$3</td>
<td>200</td>
<td>$4</td>
<td>150</td>
</tr>
</tbody>
</table>

Calculating Nominal GDP

2008 \( ($1 \text{ per hot dog} \times 100 \text{ hot dogs}) + ($2 \text{ per hamburger} \times 50 \text{ hamburgers}) = $200 \)
2009 \( ($2 \text{ per hot dog} \times 150 \text{ hot dogs}) + ($3 \text{ per hamburger} \times 100 \text{ hamburgers}) = $600 \)
2010 \( ($3 \text{ per hot dog} \times 200 \text{ hot dogs}) + ($4 \text{ per hamburger} \times 150 \text{ hamburgers}) = $1,200 \)

Calculating Real GDP (base year 2008)

2008 \( ($1 \text{ per hot dog} \times 100 \text{ hot dogs}) + ($2 \text{ per hamburger} \times 50 \text{ hamburgers}) = $200 \)
2009 \( ($1 \text{ per hot dog} \times 150 \text{ hot dogs}) + ($2 \text{ per hamburger} \times 100 \text{ hamburgers}) = $350 \)
2010 \( ($1 \text{ per hot dog} \times 200 \text{ hot dogs}) + ($2 \text{ per hamburger} \times 150 \text{ hamburgers}) = $500 \)

Calculating the GDP Deflator

2008 \( ($200 / 200) \times 100 = 100 \)
2009 \( ($600 / 350) \times 100 = 171 \)
2010 \( ($1,200 / 500) \times 100 = 240 \)
The table shows the calculation of nominal GDP for these three years. Total spending rises from $200 in 2008 to $600 in 2009 and then to $1,200 in 2010. Part of this rise is attributable to the increase in the quantities of hot dogs and hamburgers, and part is attributable to the increase in the prices of hot dogs and hamburgers.

To obtain a measure of the amount produced that is not affected by changes in prices, we use real GDP, which is the production of goods and services valued at constant prices. We calculate real GDP by first designating one year as a base year. We then use the prices of hot dogs and hamburgers in the base year to compute the value of goods and services in all the years. In other words, the prices in the base year provide the basis for comparing quantities in different years.

Suppose that we choose 2008 to be the base year in our example. We can then use the prices of hot dogs and hamburgers in 2008 to compute the value of goods and services produced in 2008, 2009, and 2010. Table 2 shows these calculations. To compute real GDP for 2008, we use the prices of hot dogs and hamburgers in 2008 (the base year) and the quantities of hot dogs and hamburgers produced in 2008. (Thus, for the base year, real GDP always equals nominal GDP.) To compute real GDP for 2009, we use the prices of hot dogs and hamburgers in 2008 (the base year) and the quantities of hot dogs and hamburgers produced in 2009. Similarly, to compute real GDP for 2010, we use the prices in 2008 and the quantities in 2010. When we find that real GDP has risen from $200 in 2008 to $350 in 2009 and then to $500 in 2010, we know that the increase is attributable to an increase in the quantities produced because the prices are being held fixed at base-year levels.

To sum up: Nominal GDP uses current prices to place a value on the economy’s production of goods and services. Real GDP uses constant base-year prices to place a value on the economy’s production of goods and services. Because real GDP is not affected by changes in prices, changes in real GDP reflect only changes in the amounts being produced. Thus, real GDP is a measure of the economy’s production of goods and services.

Our goal in computing GDP is to gauge how well the overall economy is performing. Because real GDP measures the economy’s production of goods and services, it reflects the economy’s ability to satisfy people’s needs and desires. Thus, real GDP is a better gauge of economic well-being than is nominal GDP. When economists talk about the economy’s GDP, they usually mean real GDP rather than nominal GDP. And when they talk about growth in the economy, they measure that growth as the percentage change in real GDP from one period to another.

The GDP Deflator

As we have just seen, nominal GDP reflects both the quantities of goods and services the economy is producing and the prices of those goods and services. By contrast, by holding prices constant at base-year levels, real GDP reflects only the quantities produced. From these two statistics, we can compute a third, called the GDP deflator, which reflects only the prices of goods and services.

The GDP deflator is calculated as follows:

\[
\text{GDP deflator} = \frac{\text{Nominal GDP}}{\text{Real GDP}} \times 100.
\]

Because nominal GDP and real GDP must be the same in the base year, the GDP deflator for the base year always equals 100. The GDP deflator for subsequent
years measures the change in nominal GDP from the base year that cannot be attributable to a change in real GDP.

The GDP deflator measures the current level of prices relative to the level of prices in the base year. To see why this is true, consider a couple of simple examples. First, imagine that the quantities produced in the economy rise over time but prices remain the same. In this case, both nominal and real GDP rise together, so the GDP deflator is constant. Now suppose, instead, that prices rise over time but the quantities produced stay the same. In this second case, nominal GDP rises but real GDP remains the same, so the GDP deflator rises as well. Notice that, in both cases, the GDP deflator reflects what’s happening to prices, not quantities.

Let’s now return to our numerical example in Table 2. The GDP deflator is computed at the bottom of the table. For year 2008, nominal GDP is $200, and real GDP is $200, so the GDP deflator is 100. (The deflator is always 100 in the base year.) For the year 2009, nominal GDP is $600, and real GDP is $350, so the GDP deflator is 171.

Economists use the term inflation to describe a situation in which the economy’s overall price level is rising. The inflation rate is the percentage change in some measure of the price level from one period to the next. Using the GDP deflator, the inflation rate between two consecutive years is computed as follows:

\[
\text{Inflation rate in year 2} = \frac{\text{GDP deflator in year 2} - \text{GDP deflator in year 1}}{\text{GDP deflator in year 1}} \times 100.
\]

Because the GDP deflator rose in year 2009 from 100 to 171, the inflation rate is \(100 \times (171 - 100)/100\), or 71 percent. In 2010, the GDP deflator rose to 240 from 171 the previous year, so the inflation rate is \(100 \times (240 - 171)/171\), or 40 percent.

The GDP deflator is one measure that economists use to monitor the average level of prices in the economy and thus the rate of inflation. The GDP deflator gets its name because it can be used to take inflation out of nominal GDP—that is, to “deflate” nominal GDP for the rise that is due to increases in prices. We examine another measure of the economy’s price level, called the consumer price index, in the next chapter, where we also describe the differences between the two measures.

REAL GDP OVER RECENT HISTORY

Now that we know how real GDP is defined and measured, let’s look at what this macroeconomic variable tells us about the recent history of the United States. Figure 2 shows quarterly data on real GDP for the U.S. economy since 1965.

The most obvious feature of these data is that real GDP grows over time. The real GDP of the U.S. economy in 2007 was almost four times its 1965 level. Put differently, the output of goods and services produced in the United States has grown on average 3.2 percent per year. This continued growth in real GDP enables the typical American to enjoy greater economic prosperity than his or her parents and grandparents did.

A second feature of the GDP data is that growth is not steady. The upward climb of real GDP is occasionally interrupted by periods during which GDP declines, called recessions. Figure 2 marks recessions with shaded vertical bars. (There is no ironclad rule for when the official business cycle dating committee will declare that a recession has occurred, but an old rule of thumb is two consecutive quarters...
of falling real GDP.) Recessions are associated not only with lower incomes but also with other forms of economic distress: rising unemployment, falling profits, increased bankruptcies, and so on.

Much of macroeconomics is aimed at explaining the long-run growth and short-run fluctuations in real GDP. As we will see in the coming chapters, we need different models for these two purposes. Because the short-run fluctuations represent deviations from the long-run trend, we first examine the behavior of key macroeconomic variables, including real GDP, in the long run. Then in later chapters, we build on this analysis to explain short-run fluctuations.

**QUICK QUIZ** Define real GDP and nominal GDP. Which is a better measure of economic well-being? Why?

**IS GDP A GOOD MEASURE OF ECONOMIC WELL-BEING?**

Earlier in this chapter, GDP was called the best single measure of the economic well-being of a society. Now that we know what GDP is, we can evaluate this claim.

As we have seen, GDP measures both the economy’s total income and the economy’s total expenditure on goods and services. Thus, GDP per person tells us the income and expenditure of the average person in the economy. Because most people would prefer to receive higher income and enjoy higher expenditure, GDP per person seems a natural measure of the economic well-being of the average individual.

Yet some people dispute the validity of GDP as a measure of well-being. When Senator Robert Kennedy was running for president in 1968, he gave a moving critique of such economic measures:
Gross domestic product does not allow for the health of our children, the quality of their education, or the joy of their play. It does not include the beauty of our poetry or the strength of our marriages, the intelligence of our public debate or the integrity of our public officials. It measures neither our courage, nor our wisdom, nor our devotion to our country. It measures everything, in short, except that which makes life worthwhile, and it can tell us everything about America except why we are proud that we are Americans.

Much of what Robert Kennedy said is correct. Why, then, do we care about GDP?

The answer is that a large GDP does in fact help us to lead a good life. GDP does not measure the health of our children, but nations with larger GDP can afford better healthcare for their children. GDP does not measure the quality of

---

**The Underground Economy**

The gross domestic product misses many transactions that take place in the underground economy.

**Searching for the Hidden Economy**

By Doug Campbell

Here is the brief, unremarkable story of how I recently came to participate in the underground economy:

Midafternoon on the iciest day this past winter, a man knocked at my front door. “Shovel your walk?” he asked. “Only $5.”

Outside, it was a bone-chilling 15 degrees. “Sold,” I said. A half-hour later I handed over a five-dollar bill and thanked him for saving me the trouble.

Officially, this was an unofficial transaction—off the books, with no taxes paid or safety regulations followed. (At least, I assume this hired hand didn’t bother to report that income or register with the proper authorities.) As such, it was technically illegal. And, of course, it’s the sort of thing that happens all the time.

The size of the official U.S. economy, as measured by Gross Domestic Product (GDP), was almost $12 trillion in 2004. Measurements of the unofficial economy—not including illegal activities like drug dealing and prostitution—differ substantially. But it’s generally agreed to be significant, somewhere between 6 percent and 20 percent of GDP. At the midpoint, this would be about $1.5 trillion a year.

Broadly defined, the underground, gray, informal, or shadow economy involves otherwise legal transactions that go unreported or unrecorded. That’s a wide net, capturing everything from babysitting fees, to bartering home repairs with a neighbor, to failing to report pay from moonlighting gigs. The “underground” label tends to make it sound much more sinister than it really is.

Criminal activities make up a large portion of what could be termed the total underground economy. Many studies have

<table>
<thead>
<tr>
<th>Country</th>
<th>Underground Economy as a Percentage of GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bolivia</td>
<td>68 percent</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>63</td>
</tr>
<tr>
<td>Peru</td>
<td>61</td>
</tr>
<tr>
<td>Thailand</td>
<td>54</td>
</tr>
<tr>
<td>Mexico</td>
<td>33</td>
</tr>
<tr>
<td>Argentina</td>
<td>29</td>
</tr>
<tr>
<td>Sweden</td>
<td>18</td>
</tr>
<tr>
<td>Australia</td>
<td>13</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>12</td>
</tr>
<tr>
<td>Japan</td>
<td>11</td>
</tr>
<tr>
<td>Switzerland</td>
<td>9</td>
</tr>
<tr>
<td>United States</td>
<td>8</td>
</tr>
</tbody>
</table>

Source: Friedrich Schneider. Figures are for 2002.
been done on the economics of drug dealing, prostitution, and gambling. But because money from crime is almost never recovered, many policymakers are more interested in portions of the underground economy that otherwise would be legal if not hidden from authorities. Things like shoveling walks.

Despite its intrigue, the informal economy’s importance and consequences remain in debate. The reason: “You’re trying to measure a phenomenon whose entire purpose is to hide itself from observation,” says Ed Feige, an economist at the University of Wisconsin.

This uncertainty poses problems for policymakers. Without knowing the precise size, scope, and causes of the underground economy, how can they decide what—if anything—to do about it?

Was the man who shoveled my walk engaging in a socially positive or negative activity? Was I? Suffice it to say, some economists have dedicated their entire careers to answering questions about the underground economy—and still there is nothing close to a consensus about its size or description. . . .

A SHADOWY ENTERPRISE?

Economists generally agree that the shadow economy is worse in developing nations, whose webs of bureaucratic red tape and corruption are notorious. For instance, [economist Friedrich] Schneider in 2003 published “shadow economy” estimates (defined broadly as all market-based, legal production of goods and services deliberately concealed from the authorities) for countries including: Zimbabwe, estimated at a whopping 63.2 percent of GDP, Thailand’s at 54.1 percent, and Bolivia’s at 68.3 percent. Among former Soviet bloc nations, Georgia led the way with a 68 percent of GDP shadow economy, and together those nations had an average 40.1 percent of GDP underground. This contrasts with an average of 16.7 percent among Western nations. . . .

In his 2003 book, Reofer Madness: Sex, Drugs and Cheap Labor in the American Black Market, investigative writer Eric Schlosser invokes Adam Smith’s “invisible hand” theory that men pursuing their own self-interest will generate benefits for society as a whole. This invisible hand has produced a fairly sizable underground economy, and we cannot understand our entire economic system without understanding how the hidden underbelly functions, too. “The underground is a good measure of the progress and the health of nations,” Schlosser writes. “When much is wrong, much needs to be hidden.” Schlosser’s implication was that much is wrong in the United States. If he had taken a more global view, he might have decided relatively little is hidden here.


their education, but nations with larger GDP can afford better educational systems. GDP does not measure the beauty of our poetry, but nations with larger GDP can afford to teach more of their citizens to read and enjoy poetry. GDP does not take account of our intelligence, integrity, courage, wisdom, or devotion to country, but all of these laudable attributes are easier to foster when people are less concerned about being able to afford the material necessities of life. In short, GDP does not directly measure those things that make life worthwhile, but it does measure our ability to obtain many of the inputs into a worthwhile life.

GDP is not, however, a perfect measure of well-being. Some things that contribute to a good life are left out of GDP. One is leisure. Suppose, for instance, that everyone in the economy suddenly started working every day of the week, rather than enjoying leisure on weekends. More goods and services would be produced, and GDP would rise. Yet despite the increase in GDP, we should not conclude
that everyone would be better off. The loss from reduced leisure would offset the gain from producing and consuming a greater quantity of goods and services.

Because GDP uses market prices to value goods and services, it excludes the value of almost all activity that takes place outside markets. In particular, GDP omits the value of goods and services produced at home. When a chef prepares a delicious meal and sells it at his restaurant, the value of that meal is part of GDP. But if the chef prepares the same meal for his family, the value he has added to the raw ingredients is left out of GDP. Similarly, child care provided in day-care centers is part of GDP, whereas child care by parents at home is not. Volunteer work also contributes to the well-being of those in society, but GDP does not reflect these contributions.

Another thing that GDP excludes is the quality of the environment. Imagine that the government eliminated all environmental regulations. Firms could then produce goods and services without considering the pollution they create, and GDP might rise. Yet well-being would most likely fall. The deterioration in the quality of air and water would more than offset the gains from greater production.

GDP also says nothing about the distribution of income. A society in which 100 people have annual incomes of $50,000 has GDP of $5 million and, not surprisingly, GDP per person of $50,000. So does a society in which 10 people earn $500,000 and 90 suffer with nothing at all. Few people would look at those two situations and call them equivalent. GDP per person tells us what happens to the average person, but behind the average lies a large variety of personal experiences.

In the end, we can conclude that GDP is a good measure of economic well-being for most—but not all—purposes. It is important to keep in mind what GDP includes and what it leaves out.

**Case Study: International Differences in GDP and the Quality of Life**

One way to gauge the usefulness of GDP as a measure of economic well-being is to examine international data. Rich and poor countries have vastly different levels of GDP per person. If a large GDP leads to a higher standard of living, then we should observe GDP to be strongly correlated with various measures of the quality of life. And, in fact, we do.

Table 3 shows twelve of the world’s most populous countries ranked in order of GDP per person. The table also shows life expectancy (the expected life span at birth), literacy (the percentage of the adult population who can read), and Internet usage (the percentage of the population that regularly uses the Internet). These data show a clear pattern. In rich countries, such as the United States, Japan, and Germany, people can expect to live to about 80, almost all of the population can read, and a half to two-thirds of the population uses the Internet. In poor countries, such as Nigeria, Bangladesh, and Pakistan, people typically die 10 to 20 years earlier, a substantial share of the population is illiterate, and Internet usage is rare.

Data on other aspects of the quality of life tell a similar story. Countries with low GDP per person tend to have more infants with low birth weight, higher rates of infant mortality, higher rates of maternal mortality, higher rates of child malnutrition, and less common access to safe drinking water. In countries with
low GDP per person, fewer school-age children are actually in school, and those who are in school must learn with fewer teachers per student. These countries also tend to have fewer televisions, fewer telephones, fewer paved roads, and fewer households with electricity. International data leave no doubt that a nation’s GDP per person is closely associated with its citizens’ standard of living.

QUICK QUIZ Why should policymakers care about GDP?

CONCLUSION

This chapter has discussed how economists measure the total income of a nation. Measurement is, of course, only a starting point. Much of macroeconomics is aimed at revealing the long-run and short-run determinants of a nation’s gross domestic product. Why, for example, is GDP higher in the United States and Japan than in India and Nigeria? What can the governments of the poorest countries do to promote more rapid GDP growth? Why does GDP in the United States rise rapidly in some years and fall in others? What can U.S. policymakers do to reduce the severity of these fluctuations in GDP? These are the questions we will take up shortly.

At this point, it is important to acknowledge the significance of just measuring GDP. We all get some sense of how the economy is doing as we go about our lives. But the economists who study changes in the economy and the policymakers who formulate economic policies need more than this vague sense—they need concrete data on which to base their judgments. Quantifying the behavior of the economy with statistics such as GDP is, therefore, the first step to developing a science of macroeconomics.
Who Wins at the Olympics?

Every four years, the nations of the world compete in the Olympic Games. When the games end, commentators use the number of medals a nation takes home as a measure of success. This measure seems very different from the GDP that economists use to measure success. It turns out, however, that this is not so.

Economists Andrew Bernard and Meghan Busse examined the determinants of Olympic success in a study published in the *Review of Economics and Statistics* in 2004. The most obvious explanation is population: Countries with more people will, other things equal, have more star athletes. But this is not the full story. China, India, Indonesia, and Bangladesh together have more than 40 percent of the world’s population, but they typically win only 6 percent of the medals. The reason is that these countries are poor: Despite their large populations, they account for only 5 percent of the world’s GDP. Their poverty prevents many gifted athletes from reaching their potential.

Bernard and Busse find that the best gauge of a nation’s ability to produce world-class athletes is total GDP. A large total GDP means more medals, regardless of whether the total comes from high GDP per person or a large number of people. In other words, if two nations have the same total GDP, they can be expected to win the same number of medals, even if one nation (India) has many people and low GDP per person and the other nation (Netherlands) has few people and high GDP per person.

In addition to GDP, two other factors influence the number of medals won. The host country usually earns extra medals, reflecting the benefit that athletes get from competing on their home turf. In addition, the former communist countries of Eastern Europe (the Soviet Union, Romania, East Germany, and so on) earned more medals than other countries with similar GDP. These centrally planned economies devoted more of the nation’s resources to training Olympic athletes than did free-market economies, where people have more control over their own lives.

**SUMMARY**

- Because every transaction has a buyer and a seller, the total expenditure in the economy must equal the total income in the economy.

- Gross domestic product (GDP) measures an economy’s total expenditure on newly produced goods and services and the total income earned from the production of these goods and services. More precisely, GDP is the market value of all final goods and services produced within a country in a given period of time.

- GDP is divided among four components of expenditure: consumption, investment, government purchases, and net exports. Consumption includes spending on goods and services by households, with the exception of purchases of new housing. Investment includes spending on new equipment and structures, including households’ purchases of new housing. Government purchases include spending on goods and services by local, state, and federal governments. Net exports equal the value of goods and services produced domestically and sold abroad (exports) minus the value of goods and services produced abroad and sold domestically (imports).

- Nominal GDP uses current prices to value the economy’s production of goods and services.
Real GDP uses constant base-year prices to value the economy’s production of goods and services. The GDP deflator—calculated from the ratio of nominal to real GDP—measures the level of prices in the economy.

- GDP is a good measure of economic well-being because people prefer higher to lower incomes. But it is not a perfect measure of well-being. For example, GDP excludes the value of leisure and the value of a clean environment.

### Key Concepts

- microeconomics, p. 323
- macroeconomics, p. 324
- gross domestic product (GDP), p. 326
- consumption, p. 329
- investment, p. 329
- government purchases, p. 330
- net exports, p. 330
- nominal GDP, p. 332
- real GDP, p. 333
- GDP deflator, p. 333

### Questions for Review

1. Explain why an economy’s income must equal its expenditure.
2. Which contributes more to GDP—the production of an economy car or the production of a luxury car? Why?
3. A farmer sells wheat to a baker for $2. The baker uses the wheat to make bread, which is sold for $3. What is the total contribution of these transactions to GDP?
4. Many years ago, Peggy paid $500 to put together a record collection. Today, she sold her albums at a garage sale for $100. How does this sale affect current GDP?
5. List the four components of GDP. Give an example of each.
6. Why do economists use real GDP rather than nominal GDP to gauge economic well-being?
7. In the year 2010, the economy produces 100 loaves of bread that sell for $2 each. In the year 2011, the economy produces 200 loaves of bread that sell for $3 each. Calculate nominal GDP, real GDP, and the GDP deflator for each year. (Use 2010 as the base year.) By what percentage does each of these three statistics rise from one year to the next?
8. Why is it desirable for a country to have a large GDP? Give an example of something that would raise GDP and yet be undesirable.

### Problems and Applications

1. What components of GDP (if any) would each of the following transactions affect? Explain.
   a. A family buys a new refrigerator.
   b. Aunt Jane buys a new house.
   c. Ford sells a Mustang from its inventory.
   d. You buy a pizza.
   e. California repaves Highway 101.
   f. Your parents buy a bottle of French wine.
   g. Honda expands its factory in Marysville, Ohio.
2. The government purchases component of GDP does not include spending on transfer payments such as Social Security. Thinking about the definition of GDP, explain why transfer payments are excluded.
3. As the chapter states, GDP does not include the value of used goods that are resold. Why would including such transactions make GDP a less informative measure of economic well-being?
4. Consider an economy that produces only one good. In year 1, the quantity produced is \( Q_1 \) and the price is \( P_1 \). In year 2, the quantity produced is \( Q_2 \) and the price is \( P_2 \). In year 3, the quantity produced is \( Q_3 \) and the price is \( P_3 \). Year 1 is the
base year. Answer the following questions in terms of these variables, and be sure to simplify your answer if possible.
a. What is nominal GDP for each of these three years?
b. What is real GDP for each of these years?
c. What is the GDP deflator for each of these years?
d. What is the percentage growth rate of real GDP from year 2 to year 3?
e. What is the inflation rate as measured by the GDP deflator from year 2 to year 3?

5. Below are some data from the land of milk and honey.

<table>
<thead>
<tr>
<th>Year</th>
<th>Price of Milk</th>
<th>Quantity of Milk</th>
<th>Price of Honey</th>
<th>Quantity of Honey</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>$1</td>
<td>100 quarts</td>
<td>$2</td>
<td>50 quarts</td>
</tr>
<tr>
<td>2009</td>
<td>$1</td>
<td>200</td>
<td>$2</td>
<td>100</td>
</tr>
<tr>
<td>2010</td>
<td>$2</td>
<td>200</td>
<td>$4</td>
<td>100</td>
</tr>
</tbody>
</table>

a. Compute nominal GDP, real GDP, and the GDP deflator for each year, using 2008 as the base year.
b. Compute the percentage change in nominal GDP, real GDP, and the GDP deflator in 2009 and 2010 from the preceding year. For each year, identify the variable that does not change. Explain in words why your answer makes sense.
c. Did economic well-being rise more in 2009 or 2010? Explain.

6. Consider the following data on U.S. GDP:

<table>
<thead>
<tr>
<th>Year</th>
<th>Nominal GDP (in billions of dollars)</th>
<th>GDP Deflator (base year 1996)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>9,873</td>
<td>118</td>
</tr>
<tr>
<td>1999</td>
<td>9,269</td>
<td>113</td>
</tr>
</tbody>
</table>

a. What was the growth rate of nominal GDP between 1999 and 2000? (Note: The growth rate is the percentage change from one period to the next.)
b. What was the growth rate of the GDP deflator between 1999 and 2000?
c. What was real GDP in 1999 measured in 1996 prices?
d. What was real GDP in 2000 measured in 1996 prices?
e. What was the growth rate of real GDP between 1999 and 2000?
f. Was the growth rate of nominal GDP higher or lower than the growth rate of real GDP? Explain.

7. Revised estimates of U.S. GDP are usually released by the government near the end of each month. Find a newspaper article that reports on the most recent release, or read the news release yourself at http://www.bea.gov, the website of the U.S. Bureau of Economic Analysis. Discuss the recent changes in real and nominal GDP and in the components of GDP.

8. A farmer grows wheat, which he sells to a miller for $100. The miller turns the wheat into flour, which he sells to a baker for $150. The baker turns the wheat into bread, which he sells to consumers for $180. Consumers eat the bread.

a. What is GDP in this economy? Explain.
b. Value added is defined as the value of a producer’s output minus the value of the intermediate goods that the producer buys to make the output. Assuming there are no intermediate goods beyond those described above, calculate the value added of each of the three producers.
c. What is total value added of the three producers in this economy? How does it compare to the economy’s GDP? Does this example suggest another way of calculating GDP?

9. Goods and services that are not sold in markets, such as food produced and consumed at home, are generally not included in GDP. Can you think of how this might cause the numbers in the second column of Table 3 to be misleading in a comparison of the economic well-being of the United States and India? Explain.

10. The participation of women in the U.S. labor force has risen dramatically since 1970.
a. How do you think this rise affected GDP?
b. Now imagine a measure of well-being that includes time spent working in the home and taking leisure. How would the change in this measure of well-being compare to the change in GDP?
10. In his business to add new equipment in the future. From the $220 that Barry takes home, he pays $70 in income taxes. Based on this information, compute Barry’s contribution to the following measures of income.

   a. gross domestic product
   b. net national product
   c. national income
   d. personal income
   e. disposable personal income

11. One day, Barry the Barber, Inc., collects $400 for haircuts. Over this day, his equipment depreciates in value by $50. Of the remaining $350, Barry sends $30 to the government in sales taxes, takes home $220 in wages, and retains $100 in his business to add new equipment in the future. From the $220 that Barry takes home, he pays $70 in income taxes. Based on this information, compute Barry’s contribution to the following measures of income.

   a. gross domestic product
   b. net national product
   c. national income
   d. personal income
   e. disposable personal income

   c. Can you think of other aspects of well-being that are associated with the rise in women’s labor-force participation? Would it be practical to construct a measure of well-being that includes these aspects?
Measuring the Cost of Living

In 1931, as the U.S. economy was suffering through the Great Depression, the New York Yankees paid famed baseball player Babe Ruth a salary of $80,000. At the time, this pay was extraordinary, even among the stars of baseball. According to one story, a reporter asked Ruth whether he thought it was right that he made more than President Herbert Hoover, who had a salary of only $75,000. Ruth replied, “I had a better year.”

In 2007, the median salary earned by a player on the New York Yankees was $4.8 million, and shortstop Alex Rodriguez was paid $28 million. At first, this fact might lead you to think that baseball has become vastly more lucrative over the past seven decades. But as everyone knows, the prices of goods and services have also risen. In 1931, a nickel would buy an ice-cream cone, and a quarter would buy a ticket at the local movie theater. Because prices were so much lower in Babe Ruth’s day than they are today, it is not clear whether Ruth enjoyed a higher or lower standard of living than today’s players.

In the preceding chapter, we looked at how economists use gross domestic product (GDP) to measure the quantity of goods and services that the economy is producing. This chapter examines how economists measure the overall cost of living. To compare Babe Ruth’s salary of $80,000 to salaries from today, we need to find some way of turning dollar figures into meaningful measures of purchasing power. That is exactly the job of a statistic called the consumer price index. After
seeing how the consumer price index is constructed, we discuss how we can use such a price index to compare dollar figures from different points in time.

The consumer price index is used to monitor changes in the cost of living over time. When the consumer price index rises, the typical family has to spend more money to maintain the same standard of living. Economists use the term inflation to describe a situation in which the economy’s overall price level is rising. The inflation rate is the percentage change in the price level from the previous period. The preceding chapter showed how economists can measure inflation using the GDP deflator. The inflation rate you are likely to hear on the nightly news, however, is not calculated from this statistic. Because the consumer price index better reflects the goods and services bought by consumers, it is the more common gauge of inflation.

As we will see in the coming chapters, inflation is a closely watched aspect of macroeconomic performance and is a key variable guiding macroeconomic policy. This chapter provides the background for that analysis by showing how economists measure the inflation rate using the consumer price index and how this statistic can be used to compare dollar figures from different times.

THE CONSUMER PRICE INDEX

The consumer price index (CPI) is a measure of the overall cost of the goods and services bought by a typical consumer. Each month, the Bureau of Labor Statistics (BLS), which is part of the Department of Labor, computes and reports the consumer price index. In this section, we discuss how the consumer price index is calculated and what problems arise in its measurement. We also consider how this index compares to the GDP deflator, another measure of the overall level of prices, which we examined in the preceding chapter.

HOW THE CONSUMER PRICE INDEX IS CALCULATED

When the Bureau of Labor Statistics calculates the consumer price index and the inflation rate, it uses data on the prices of thousands of goods and services. To see exactly how these statistics are constructed, let’s consider a simple economy in which consumers buy only two goods: hot dogs and hamburgers. Table 1 shows the five steps that the BLS follows.

1. Fix the basket. Determine which prices are most important to the typical consumer. If the typical consumer buys more hot dogs than hamburgers, then the price of hot dogs is more important than the price of hamburgers and, therefore, should be given greater weight in measuring the cost of living. The Bureau of Labor Statistics sets these weights by surveying consumers to find the basket of goods and services bought by the typical consumer. In the example in the table, the typical consumer buys a basket of 4 hot dogs and 2 hamburgers.

2. Find the prices. Find the prices of each of the goods and services in the basket at each point in time. The table shows the prices of hot dogs and hamburgers for 3 different years.

3. Compute the basket’s cost. Use the data on prices to calculate the cost of the basket of goods and services at different times. The table shows this calculation for each of the 3 years. Notice that only the prices in this calculation...
change. By keeping the basket of goods the same (4 hot dogs and 2 hamburgers), we are isolating the effects of price changes from the effect of any quantity changes that might be occurring at the same time.

4. Choose a base year and compute the index. Designate one year as the base year, the benchmark against which other years are compared. (The choice of base year is arbitrary, as the index is used to measure changes in the cost of living.) Once the base year is chosen, the index is calculated as follows:

\[
\text{Consumer price index} = \frac{\text{Price of basket of goods and services in current year}}{\text{Price of basket in base year}} \times 100.
\]

That is, the price of the basket of goods and services in each year is divided by the price of the basket in the base year, and this ratio is then multiplied by 100. The resulting number is the consumer price index.

In the example in the table, 2008 is the base year. In this year, the basket of hot dogs and hamburgers costs $8. Therefore, the price of the basket in all
years is divided by $8 and multiplied by 100. The consumer price index is 100 in 2008. (The index is always 100 in the base year.) The consumer price index is 175 in 2009. This means that the price of the basket in 2009 is 175 percent of its price in the base year. Put differently, a basket of goods that costs $100 in the base year costs $175 in 2009. Similarly, the consumer price index is 250 in 2010, indicating that the price level in 2010 is 250 percent of the price level in the base year.

5. **Compute the inflation rate.** Use the consumer price index to calculate the **inflation rate**, which is the percentage change in the price index from the preceding period. That is, the inflation rate between two consecutive years is computed as follows:

\[
\text{Inflation rate in year 2} = \frac{\text{CPI in year 2} - \text{CPI in year 1}}{\text{CPI in year 1}} \times 100.
\]

As shown at the bottom of Table 1, the inflation rate in our example is 75 percent in 2009 and 43 percent in 2010.

Although this example simplifies the real world by including only two goods, it shows how the Bureau of Labor Statistics computes the consumer price index and the inflation rate. The BLS collects and processes data on the prices of thousands of goods and services every month and, by following the five foregoing steps, determines how quickly the cost of living for the typical consumer is rising. When the BLS makes its monthly announcement of the consumer price index, you can usually hear the number on the evening television news or see it in the next day’s newspaper.

In addition to the consumer price index for the overall economy, the BLS calculates several other price indexes. It reports the index for specific metropolitan areas within the country (such as Boston, New York, and Los Angeles) and for some narrow categories of goods and services (such as food, clothing, and energy). It also calculates the **producer price index** (PPI), which measures the cost of a basket of goods and services bought by firms rather than consumers. Because firms eventually pass on their costs to consumers in the form of higher consumer prices, changes in the producer price index are often thought to be useful in predicting changes in the consumer price index.

**Problems in Measuring the Cost of Living**

The goal of the consumer price index is to measure changes in the cost of living. In other words, the consumer price index tries to gauge how much incomes must rise to maintain a constant standard of living. The consumer price index, however, is not a perfect measure of the cost of living. Three problems with the index are widely acknowledged but difficult to solve.

The first problem is called **substitution bias**. When prices change from one year to the next, they do not all change proportionately: Some prices rise more than others. Consumers respond to these differing price changes by buying less of the goods whose prices have risen by relatively large amounts and by buying more of the goods whose prices have risen less or perhaps even have fallen. That is, consumers substitute toward goods that have become relatively less expensive. If a price index is computed assuming a fixed basket of goods, it ignores the possibility of consumer substitution and, therefore, overstates the increase in the cost of living from one year to the next.
Let’s consider a simple example. Imagine that in the base year, apples are cheaper than pears, and so consumers buy more apples than pears. When the Bureau of Labor Statistics constructs the basket of goods, it will include more apples than pears. Suppose that next year pears are cheaper than apples. Consumers will naturally respond to the price changes by buying more pears and fewer apples. Yet when computing the consumer price index, the BLS uses a fixed basket, which in essence assumes that consumers continue buying the now expensive apples in the same quantities as before. For this reason, the index will measure a much larger increase in the cost of living than consumers actually experience.

The second problem with the consumer price index is the introduction of new goods. When a new good is introduced, consumers have more variety from which to choose, and this in turn reduces the cost of maintaining the same level of economic well-being. To see why, consider a hypothetical situation: Suppose you could choose between a $100 gift certificate at a large store that offered a wide array of goods and a $100 gift certificate at a small store with the same prices but

FIGURE 1

The Typical Basket of Goods and Services

This figure shows how the typical consumer divides spending among various categories of goods and services. The Bureau of Labor Statistics calls each percentage the “relative importance” of the category.

a more limited selection. Which would you prefer? Most people would pick the store with greater variety. In essence, the increased set of possible choices makes each dollar more valuable. The same is true with the evolution of the economy over time: As new goods are introduced, consumers have more choices, and each dollar is worth more. Yet because the consumer price index is based on a fixed basket of goods and services, it does not reflect the increase in the value of the dollar that arises from the introduction of new goods.

Again, let’s consider an example. When video cassette recorders (VCRs) were introduced in the late 1970s, consumers were able to watch their favorite movies at home. Although not a perfect substitute for a first-run movie on a large screen, an old movie in the comfort of your family room was a new option that increased consumers’ set of opportunities. For any given number of dollars, the introduction of the VCR made people better off; conversely, to achieve the same level of economic well-being required a smaller number of dollars. A perfect cost-of-living index would have reflected the introduction of the VCR with a decrease in the cost of living. The consumer price index, however, did not decrease in response to the introduction of the VCR. Eventually, the Bureau of Labor Statistics did revise the basket of goods to include VCRs, and subsequently, the index reflected changes.
magnifies price increases rather than suppressing them.

The bureau says hedonics actually helps boost the housing component of the CPI. In order to take into account the aging of housing, and presumably falling quality that goes with it, the CPI applies a form of hedonics that links the age of a housing unit to rents. If someone is paying the equivalent of $500 a month in rent for several years, the rent has actually gone up as the unit ages and becomes less desirable, according to the government.

The hub of this effort is a warren of beige-walled cubicles at the Bureau of Labor Statistics a few blocks from the Capitol. Here 40 commodity specialists hunch over reports with 85,000 price quotes that flow in from around the country every month. The numbers are gathered by 400 part-time data collectors. They visit stores and note prices on the items that make up the basket of goods in the CPI, ranging from ladies’ shoes to skim milk to microwave ovens.

There is still much debate among economists about how severe these measurement problems are and what should be done about them. Several studies written during the 1990s concluded that the consumer price index overstated inflation by about 1 percentage point per year. In response to this criticism, the Bureau of Labor Statistics adopted several technical changes to improve the CPI, and many
economists believe the bias is now only about half as large as it once was. The issue is important because many government programs use the consumer price index to adjust for changes in the overall level of prices. Recipients of Social Security, for instance, get annual increases in benefits that are tied to the consumer price index. Some economists have suggested modifying these programs to correct the measurement problems by, for instance, reducing the magnitude of the automatic benefit increases.

**The GDP Deflator versus the Consumer Price Index**

In the preceding chapter, we examined another measure of the overall level of prices in the economy—the GDP deflator. The GDP deflator is the ratio of nominal GDP to real GDP. Because nominal GDP is current output valued at current prices and real GDP is current output valued at base-year prices, the GDP deflator reflects the current level of prices relative to the level of prices in the base year.

Economists and policymakers monitor both the GDP deflator and the consumer price index to gauge how quickly prices are rising. Usually, these two statistics tell a similar story. Yet two important differences can cause them to diverge.

The first difference is that the GDP deflator reflects the prices of all goods and services produced domestically, whereas the consumer price index reflects the prices of all goods and services bought by consumers. For example, suppose that the price of an airplane produced by Boeing and sold to the Air Force rises. Even though the plane is part of GDP, it is not part of the basket of goods and services bought by a typical consumer. Thus, the price increase shows up in the GDP deflator but not in the consumer price index.

As another example, suppose that Volvo raises the price of its cars. Because Volvos are made in Sweden, the car is not part of U.S. GDP. But U.S. consumers buy Volvos, and so the car is part of the typical consumer’s basket of goods. Hence, a price increase in an imported consumption good, such as a Volvo, shows up in the consumer price index but not in the GDP deflator.

This first difference between the consumer price index and the GDP deflator is particularly important when the price of oil changes. Although the United States does produce some oil, much of the oil we use is imported. As a result, oil and oil products such as gasoline and heating oil are a much larger share of consumer spending than of GDP. When the price of oil rises, the consumer price index rises by much more than does the GDP deflator.

The second and subtler difference between the GDP deflator and the consumer price index concerns how various prices are weighted to yield a single number for the overall level of prices. The consumer price index compares the price of a fixed basket of goods and services to the price of the basket in the base year. Only occasionally does the Bureau of Labor Statistics change the basket of goods. By contrast, the GDP deflator compares the price of currently produced goods and services to the price of the same goods and services in the base year. Thus, the group of goods and services used to compute the GDP deflator changes automatically over time. This difference is not important when all prices are changing proportionately. But if the prices of different goods and services are changing by varying amounts, the way we weight the various prices matters for the overall inflation rate.
Figure 2 shows the inflation rate as measured by both the GDP deflator and the consumer price index for each year since 1965. You can see that sometimes the two measures diverge. When they do diverge, it is possible to go behind these numbers and explain the divergence with the two differences we have discussed. For example, in 1979 and 1980, CPI inflation spiked up more than the GDP deflator largely because oil prices more than doubled during these two years. Yet divergence between these two measures is the exception rather than the rule. In the 1970s, both the GDP deflator and the consumer price index show high rates of inflation. In the late 1980s, 1990s, and the first decade of the 2000s, both measures show low rates of inflation.

QUICK QUIZ  Explain briefly what the consumer price index measures and how it is constructed.

CORRECTING ECONOMIC VARIABLES FOR THE EFFECTS OF INFLATION

The purpose of measuring the overall level of prices in the economy is to permit comparison between dollar figures from different times. Now that we know how price indexes are calculated, let’s see how we might use such an index to compare a dollar figure from the past to a dollar figure in the present.
DOLLAR FIGURES FROM DIFFERENT TIMES

We first return to the issue of Babe Ruth’s salary. Was his salary of $80,000 in 1931 high or low compared to the salaries of today’s players?

To answer this question, we need to know the level of prices in 1931 and the level of prices today. Part of the increase in baseball salaries compensates players for higher prices today. To compare Ruth’s salary to those of today’s players, we need to inflate Ruth’s salary to turn 1931 dollars into today’s dollars.

The formula for turning dollar figures from year $T$ into today’s dollars is the following:

$$\text{Amount in today’s dollars} = \text{Amount in year } T \text{ dollars} \times \frac{\text{Price level today}}{\text{Price level in year } T}.$$  

A price index such as the consumer price index measures the price level and thus determines the size of the inflation correction.

Let’s apply this formula to Ruth’s salary. Government statistics show a consumer price index of 15.2 for 1931 and 207 for 2007. Thus, the overall level of prices has risen by a factor of 13.6 (which equals 207/15.2). We can use these numbers to measure Ruth’s salary in 2007 dollars, as follows:

$$\text{Salary in 2007 dollars} = \text{Salary in 1931 dollars} \times \frac{\text{Price level in 2007}}{\text{Price level in 1931}}$$

$$= \$80,000 \times \frac{207}{15.2}$$

$$= \$1,089,474$$

We find that Babe Ruth’s 1931 salary is equivalent to a salary today of over $1 million. That is a good income, but it is less than a quarter of the median Yankee salary today and only 4 percent of what the Yankees pay A-Rod. Various forces, including overall economic growth and the increasing income shares earned by superstars, have substantially raised the living standards of the best athletes.

Let’s also examine President Hoover’s 1931 salary of $75,000. To translate that figure into 2007 dollars, we again multiply the ratio of the price levels in the 2 years. We find that Hoover’s salary is equivalent to $75,000 $\times$ (207/15.2), or $1,021,382, in 2007 dollars. This is well above President George W. Bush’s salary of $400,000. It seems that President Hoover did have a pretty good year after all.

INDEXATION

As we have just seen, price indexes are used to correct for the effects of inflation when comparing dollar figures from different times. This type of correction shows up in many places in the economy. When some dollar amount is automatically corrected for changes in the price level by law or contract, the amount is said to be indexed for inflation.

For example, many long-term contracts between firms and unions include partial or complete indexation of the wage to the consumer price index. Such a provision is called a cost-of-living allowance, or COLA. A COLA automatically raises the wage when the consumer price index rises.
Indexation is also a feature of many laws. Social Security benefits, for example, are adjusted every year to compensate the elderly for increases in prices. The brackets of the federal income tax—the income levels at which the tax rates change—are also indexed for inflation. There are, however, many ways in which the tax system is not indexed for inflation, even when perhaps it should be. We discuss these issues more fully when we discuss the costs of inflation later in this book.

**REAL AND NOMINAL INTEREST RATES**

Correcting economic variables for the effects of inflation is particularly important, and somewhat tricky, when we look at data on interest rates. The very concept of an interest rate necessarily involves comparing amounts of money at different points in time. When you deposit your savings in a bank account, you give the bank some money now, and the bank returns your deposit with interest in the future. Similarly, when you borrow from a bank, you get some money now, but you will have to repay the loan with interest in the future. In both cases, to fully understand the deal between you and the bank, it is crucial to acknowledge that future dollars could have a different value than today’s dollars. That is, you have to correct for the effects of inflation.

Let’s consider an example. Suppose Sally Saver deposits $1,000 in a bank account that pays an annual interest rate of 10 percent. A year later, after Sally has accumulated $100 in interest, she withdraws her $1,100. Is Sally $100 richer than she was when she made the deposit a year earlier?
The answer depends on what we mean by “richer.” Sally does have $100 more than she had before. In other words, the number of dollars in her possession has risen by 10 percent. But Sally does not care about the amount of money itself: She cares about what she can buy with it. If prices have risen while her money was in the bank, each dollar now buys less than it did a year ago. In this case, her purchasing power—the amount of goods and services she can buy—has not risen by 10 percent.

To keep things simple, let’s suppose that Sally is a music fan and buys only music CDs. When Sally made her deposit, a CD at her local music store cost $10. Her deposit of $1,000 was equivalent to 100 CDs. A year later, after getting her 10 percent interest, she has $1,100. How many CDs can she buy now? It depends on what has happened to the price of a CD. Here are some examples:

- **Zero inflation**: If the price of a CD remains at $10, the amount she can buy has risen from 100 to 110 CDs. The 10 percent increase in the number of dollars means a 10 percent increase in her purchasing power.
- **Six percent inflation**: If the price of a CD rises from $10 to $10.60, then the number of CDs she can buy has risen from 100 to approximately 104. Her purchasing power has increased by about 4 percent.
- **Ten percent inflation**: If the price of a CD rises from $10 to $11, she can still buy only 100 CDs. Even though Sally’s dollar wealth has risen, her purchasing power is the same as it was a year earlier.
- **Twelve percent inflation**: If the price of a CD increases from $10 to $11.20, the number of CDs she can buy has fallen from 100 to approximately 98. Even with her greater number of dollars, her purchasing power has decreased by about 2 percent.

And if Sally were living in an economy with deflation—falling prices—another possibility could arise:

- **Two percent deflation**: If the price of a CD falls from $10 to $9.80, then the number of CDs she can buy rises from 100 to approximately 112. Her purchasing power increases by about 12 percent.

These examples show that the higher the rate of inflation, the smaller the increase in Sally’s purchasing power. If the rate of inflation exceeds the rate of interest, her purchasing power actually falls. And if there is deflation (that is, a negative rate of inflation), her purchasing power rises by more than the rate of interest.

To understand how much a person earns in a savings account, we need to consider both the interest rate and the change in the prices. The interest rate that measures the change in dollar amounts is called the **nominal interest rate**, and the interest rate corrected for inflation is called the **real interest rate**. The nominal interest rate, the real interest rate, and inflation are related approximately as follows:

\[
\text{Real interest rate} = \text{Nominal interest rate} - \text{Inflation rate}.
\]

The real interest rate is the difference between the nominal interest rate and the rate of inflation. The nominal interest rate tells you how fast the number of dollars in your bank account rises over time, while the real interest rate tells you how fast the purchasing power of your bank account rises over time.
INTEREST RATES IN THE U.S. ECONOMY

Figure 3 shows real and nominal interest rates in the U.S. economy since 1965. The nominal interest rate in this figure is the rate on 3-month Treasury bills (although data on other interest rates would be similar). The real interest rate is computed by subtracting the rate of inflation from this nominal interest rate. Here the inflation rate is measured as the percentage change in the consumer price index.

One feature of this figure is that the nominal interest rate always exceeds the real interest rate. This reflects the fact that the U.S. economy has experienced rising consumer prices in every year during this period. By contrast, if you look at data for the U.S. economy during the late 19th century or for the Japanese economy in some recent years, you will find periods of deflation. During deflation, the real interest rate exceeds the nominal interest rate.

The figure also shows that because inflation is variable, real and nominal interest rates do not always move together. For example, in the late 1970s, nominal interest rates were high. But because inflation was very high, real interest rates were low. Indeed, during much of the 1970s, real interest rates were negative, for inflation eroded people’s savings more quickly than nominal interest payments increased them. By contrast, in the late 1990s, nominal interest rates were lower than they had been two decades earlier. But because inflation was much lower, real interest rates were higher. In the coming chapters, we will examine the economic forces that determine both real and nominal interest rates.

QUICK QUIZ  Henry Ford paid his workers $5 a day in 1914. If the consumer price index was 10 in 1914 and 207 in 2007, how much is the Ford paycheck worth in 2007 dollars?
“A nickel ain’t worth a dime anymore,” baseball player Yogi Berra once observed. Indeed, throughout recent history, the real values behind the nickel, the dime, and the dollar have not been stable. Persistent increases in the overall level of prices have been the norm. Such inflation reduces the purchasing power of each unit of money over time. When comparing dollar figures from different times, it is important to keep in mind that a dollar today is not the same as a dollar 20 years ago or, most likely, 20 years from now.

This chapter has discussed how economists measure the overall level of prices in the economy and how they use price indexes to correct economic variables for the effects of inflation. Price indexes allow us to compare dollar figures from different points in time and, therefore, get a better sense of how the economy is changing.

The discussion of price indexes in this chapter, together with the preceding chapter’s discussion of GDP, is only a first step in the study of macroeconomics. We have not yet examined what determines a nation’s GDP or the causes and effects of inflation. To do that, we need to go beyond issues of measurement. Indeed, that is our next task. Having explained how economists measure macroeconomic quantities and prices in the past two chapters, we are now ready to develop the models that explain movements in these variables.

Here is our strategy in the upcoming chapters. First, we look at the long-run determinants of real GDP and related variables, such as saving, investment, real interest rates, and unemployment. Second, we look at the long-run determinants of the price level and related variables, such as the money supply, inflation, and nominal interest rates. Last of all, having seen how these variables are determined in the long run, we examine the more complex question of what causes short-run fluctuations in real GDP and the price level. In all of these chapters, the measurement issues we have just discussed will provide the foundation for the analysis.

**SUMMARY**

- The consumer price index shows the cost of a basket of goods and services relative to the cost of the same basket in the base year. The index is used to measure the overall level of prices in the economy. The percentage change in the consumer price index measures the inflation rate.

- The consumer price index is an imperfect measure of the cost of living for three reasons. First, it does not take into account consumers’ ability to substitute toward goods that become relatively cheaper over time. Second, it does not take into account increases in the purchasing power of the dollar due to the introduction of new goods. Third, it is distorted by unmeasured changes in the quality of goods and services. Because of these measurement problems, the CPI overstates true inflation.

- Like the consumer price index, the GDP deflator measures the overall level of prices in the economy. Although the two price indexes usually move together, there are important differences. The GDP deflator differs from the CPI because
It includes goods and services produced rather than goods and services consumed. As a result, imported goods affect the consumer price index but not the GDP deflator. In addition, while the consumer price index uses a fixed basket of goods, the GDP deflator automatically changes the group of goods and services over time as the composition of GDP changes.

- Dollar figures from different times do not represent a valid comparison of purchasing power. To compare a dollar figure from the past to a dollar figure today, the older figure should be inflated using a price index.

Various laws and private contracts use price indexes to correct for the effects of inflation. The tax laws, however, are only partially indexed for inflation.

A correction for inflation is especially important when looking at data on interest rates. The nominal interest rate is the interest rate usually reported; it is the rate at which the number of dollars in a savings account increases over time. By contrast, the real interest rate takes into account changes in the value of the dollar over time. The real interest rate equals the nominal interest rate minus the rate of inflation.

**KEY CONCEPTS**

- consumer price index (CPI), p. 346
- producer price index, p. 348
- inflation rate, p. 348
- indexation, p. 354
- nominal interest rate, p. 356
- real interest rate, p. 356

**QUESTIONS FOR REVIEW**

1. Which do you think has a greater effect on the consumer price index: a 10 percent increase in the price of chicken or a 10 percent increase in the price of caviar? Why?
2. Describe the three problems that make the consumer price index an imperfect measure of the cost of living.
3. If the price of a Navy submarine rises, is the consumer price index or the GDP deflator affected more? Why?
4. Over a long period of time, the price of a candy bar rose from $0.10 to $0.60. Over the same period, the consumer price index rose from 150 to 300. Adjusted for overall inflation, how much did the price of the candy bar change?
5. Explain the meaning of nominal interest rate and real interest rate. How are they related?

**PROBLEMS AND APPLICATIONS**

1. Suppose that the residents of Vegopia spend all of their income on cauliflower, broccoli, and carrots. In 2008, they buy 100 heads of cauliflower for $200, 50 bunches of broccoli for $75, and 500 carrots for $50. In 2009, they buy 75 heads of cauliflower for $225, 80 bunches of broccoli for $120, and 500 carrots for $100.
   a. Calculate the price of each vegetable in each year.
   b. Using 2008 as the base year, calculate the CPI for each year.
   c. What is the inflation rate in 2009?

2. Go to the website of the Bureau of Labor Statistics (http://www.bls.gov) and find data on the consumer price index. By how much has the index including all items risen over the past year? For which categories of spending have prices risen the most? The least? Have any
5. Beginning in 1994, environmental regulations have required that gasoline contain a new additive to reduce air pollution. This requirement raised the cost of gasoline. The Bureau of Labor Statistics decided that this increase in cost represented an improvement in quality.

a. Given this decision, did the increased cost of gasoline raise the CPI?

b. What is the argument in favor of the BLS’s decision? What is the argument for a different decision?

6. Which of the problems in the construction of the CPI might be illustrated by each of the following situations? Explain.

a. the invention of the iPod
b. the introduction of air bags in cars
c. increased personal computer purchases in response to a decline in their price
d. more scoops of raisins in each package of Raisin Bran
e. greater use of fuel-efficient cars after gasoline prices increase


a. By what percentage did the price of a newspaper rise?
b. By what percentage did the wage rise?
c. In each year, how many minutes does a worker have to work to earn enough to buy a newspaper?
d. Did workers’ purchasing power in terms of newspapers rise or fall?

8. The chapter explains that Social Security benefits are increased each year in proportion to the increase in the CPI, even though most economists believe that the CPI overstates actual inflation.

a. If the elderly consume the same market basket as other people, does Social Security provide the elderly with an improvement in their standard of living each year? Explain.
b. In fact, the elderly consume more healthcare than younger people, and healthcare costs have risen faster than overall inflation. What
would you do to determine whether the elderly are actually better off from year to year?

9. When deciding how much of their income to save for retirement, should workers consider the real or the nominal interest rate that their savings will earn? Explain.

10. Suppose that a borrower and a lender agree on the nominal interest rate to be paid on a loan. Then inflation turns out to be higher than they both expected.

a. Is the real interest rate on this loan higher or lower than expected?

b. Does the lender gain or lose from this unexpectedly high inflation? Does the borrower gain or lose?

c. Inflation during the 1970s was much higher than most people had expected when the decade began. How did this affect homeowners who obtained fixed-rate mortgages during the 1960s? How did it affect the banks that lent the money?
This page intentionally left blank
PART VII

The Real Economy in the Long Run
This page intentionally left blank
Production and Growth

When you travel around the world, you see tremendous variation in the standard of living. The average income in a rich country, such as the United States, Japan, or Germany, is more than ten times the average income in a poor country, such as India, Indonesia, or Nigeria. These large differences in income are reflected in large differences in the quality of life. People in richer countries have better nutrition, safer housing, better healthcare, and longer life expectancy as well as more automobiles, more telephones, and more televisions.

Even within a country, there are large changes in the standard of living over time. In the United States over the past century, average income as measured by real GDP per person has grown by about 2 percent per year. Although 2 percent might seem small, this rate of growth implies that average income doubles every 35 years. Because of this growth, average income today is about eight times the average income a century ago. As a result, the typical American enjoys much greater economic prosperity than did his or her parents, grandparents, and great-grandparents.

Growth rates vary substantially from country to country. In recent history, some East Asian countries, such as Singapore, South Korea, and Taiwan, have experienced economic growth of about 7 percent per year. At this rate, average income doubles every 10 years. A country experiencing such rapid growth can,
in one generation, go from being among the poorest in the world to being among the richest. By contrast, in some African countries, such as Chad, Ethiopia, and Nigeria, average income has been stagnant for many years.

What explains these diverse experiences? How can rich countries maintain their high standard of living? What policies should poor countries pursue to promote more rapid growth and join the developed world? These are among the most important questions in macroeconomics. As economist Robert Lucas put it, “The consequences for human welfare in questions like these are simply staggering: Once one starts to think about them, it is hard to think about anything else.”

In the previous two chapters, we discussed how economists measure macroeconomic quantities and prices. We can now begin to study the forces that determine these variables. As we have seen, an economy’s gross domestic product (GDP) measures both the total income earned in the economy and the total expenditure on the economy’s output of goods and services. The level of real GDP is a good gauge of economic prosperity, and the growth of real GDP is a good gauge of economic progress. In this chapter we focus on the long-run determinants of the level and growth of real GDP. Later in this book, we study the short-run fluctuations of real GDP around its long-run trend.

We proceed here in three steps. First, we examine international data on real GDP per person. These data will give you some sense of how much the level and growth of living standards vary around the world. Second, we examine the role of productivity—the amount of goods and services produced for each hour of a worker’s time. In particular, we see that a nation’s standard of living is determined by the productivity of its workers, and we consider the factors that determine a nation’s productivity. Third, we consider the link between productivity and the economic policies that a nation pursues.

ECONOMIC GROWTH AROUND THE WORLD

As a starting point for our study of long-run growth, let’s look at the experiences of some of the world’s economies. Table 1 shows data on real GDP per person for thirteen countries. For each country, the data cover more than a century of history. The first and second columns of the table present the countries and time periods. (The time periods differ somewhat from country to country because of differences in data availability.) The third and fourth columns show estimates of real GDP per person about a century ago and for a recent year.

The data on real GDP per person show that living standards vary widely from country to country. Income per person in the United States, for instance, is about six times that in China and about twelve times that in India. The poorest countries have average levels of income not seen in the developed world for many decades. The typical citizen of India in 2006 had less real income than the typical resident of England in 1870. The typical person in Bangladesh in 2006 had about two-thirds the real income of a typical American a century ago.

The last column of the table shows each country’s growth rate. The growth rate measures how rapidly real GDP per person grew in the typical year. In the United States, for example, where real GDP per person was $3,752 in 1870 and $44,260 in 2006, the growth rate was 1.83 percent per year. This means that if real GDP per person, beginning at $3,752, were to increase by 1.83 percent for each of 136 years, it would end up at $44,260. Of course, real GDP per person did not actually rise
exactly 1.83 percent every year: Some years it rose by more, other years it rose by less, and in still other years it fell. The growth rate of 1.83 percent per year ignores short-run fluctuations around the long-run trend and represents an average rate of growth for real GDP per person over many years.

The countries in Table 1 are ordered by their growth rate from the most to the least rapid. Japan tops the list, with a growth rate of 2.76 percent per year. A hundred years ago, Japan was not a rich country. Japan’s average income was only somewhat higher than Mexico’s, and it was well behind Argentina’s. The standard of living in Japan in 1890 was less than half of that in India today. But because of its spectacular growth, Japan is now an economic superpower, with average income more than twice that of Mexico and Argentina and similar to Germany, Canada, and the United Kingdom. At the bottom of the list of countries are Bangladesh and Pakistan, which have experienced growth of less than 1.4 percent per year over the past century. As a result, the typical resident of these countries continues to live in abject poverty.

Because of differences in growth rates, the ranking of countries by income changes substantially over time. As we have seen, Japan is a country that has risen relative to others. One country that has fallen behind is the United Kingdom. In 1870, the United Kingdom was the richest country in the world, with average income about 20 percent higher than that of the United States and more than twice Canada’s. Today, average income in the United Kingdom is 20 percent below that of the United States and similar to Canada’s.

These data show that the world’s richest countries have no guarantee they will stay the richest and that the world’s poorest countries are not doomed forever to remain in poverty. But what explains these changes over time? Why do some countries zoom ahead while others lag behind? These are precisely the questions that we take up next.

<table>
<thead>
<tr>
<th>Country</th>
<th>Period</th>
<th>Real GDP per Person at Beginning of Period</th>
<th>Real GDP per Person at End of Period</th>
<th>Growth Rate (per year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>1890–2006</td>
<td>$1,408</td>
<td>$33,150</td>
<td>2.76%</td>
</tr>
<tr>
<td>Brazil</td>
<td>1900–2006</td>
<td>729</td>
<td>8,880</td>
<td>2.39</td>
</tr>
<tr>
<td>China</td>
<td>1900–2006</td>
<td>670</td>
<td>7,740</td>
<td>2.34</td>
</tr>
<tr>
<td>Mexico</td>
<td>1900–2006</td>
<td>1,085</td>
<td>11,410</td>
<td>2.24</td>
</tr>
<tr>
<td>Germany</td>
<td>1870–2006</td>
<td>2,045</td>
<td>31,830</td>
<td>2.04</td>
</tr>
<tr>
<td>Canada</td>
<td>1870–2006</td>
<td>2,224</td>
<td>34,610</td>
<td>2.04</td>
</tr>
<tr>
<td>Argentina</td>
<td>1900–2006</td>
<td>2,147</td>
<td>15,390</td>
<td>1.88</td>
</tr>
<tr>
<td>United States</td>
<td>1870–2006</td>
<td>3,752</td>
<td>44,260</td>
<td>1.83</td>
</tr>
<tr>
<td>India</td>
<td>1900–2006</td>
<td>632</td>
<td>3,800</td>
<td>1.71</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1870–2006</td>
<td>4,502</td>
<td>35,580</td>
<td>1.53</td>
</tr>
<tr>
<td>Indonesia</td>
<td>1900–2006</td>
<td>834</td>
<td>3,950</td>
<td>1.48</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>1900–2006</td>
<td>583</td>
<td>2,340</td>
<td>1.32</td>
</tr>
<tr>
<td>Pakistan</td>
<td>1900–2006</td>
<td>690</td>
<td>2,500</td>
<td>1.22</td>
</tr>
</tbody>
</table>

*Real GDP is measured in 2006 dollars.
George Bernard Shaw once said, “The sign of a truly educated man is to be deeply moved by statistics.” Most of us, however, have trouble being deeply moved by data on GDP—until we see what these statistics represent.

The three photos on these pages show a typical family from each of three countries—the United Kingdom, Mexico, and Mali. Each family was photographed outside their home, together with all their material possessions.

These nations have very different standards of living, as judged by these photos, GDP, or other statistics.

- The United Kingdom is an advanced economy. In 2006, its GDP per person was $35,580. A negligible share of the population lives in extreme poverty, defined here as less than $2 a day. Educational attainment is high: Among children of high school age, 95 percent are in school. Residents of the United Kingdom can expect to enjoy a long life: The probability of a person surviving to age 65 is 84 percent for men and 90 percent for women.

- Mexico is a middle-income country. In 2006, its GDP per person was $11,410. About an eighth of the population lives on less than $2 a day. Among children of high school age, 65 percent are in school. The probability of a person surviving to age 65 is 76 percent for men and 84 percent for women.

- Mali is a poor country. In 2006, its GDP per person was only $1,130. Extreme poverty is the norm: More than half of the population lives on less than $2 per day. Educational attainment in Mali is low: Among children of high school age, less than 10 percent are in school. And life is often cut short: The probability of a person surviving to age 65 is only 44 percent for men and 54 percent for women.

Economists who study economic growth try to understand what causes such large differences in the standard of living.
QUICK QUIZ
What is the approximate growth rate of real GDP per person in the United States? Name a country that has had faster growth and a country that has had slower growth.

Are You Richer Than the Richest American?

American Heritage magazine once published a list of the richest Americans of all time. The number 1 spot went to John D. Rockefeller, the oil entrepreneur who lived from 1839 to 1937. According to the magazine’s calculations, his wealth would today be the equivalent of $200 billion, more than twice that of Warren Buffett, the investor who is today’s richest American.

Despite his great wealth, Rockefeller did not enjoy many of the conveniences that we now take for granted. He couldn’t watch television, play video games, surf the Internet, or send e-mail. During the heat of summer, he couldn’t cool his home with air conditioning. For much of his life, he couldn’t travel by car or plane, and he couldn’t use a telephone to call friends or family. If he became ill, he couldn’t take advantage of many medicines, such as antibiotics, that doctors today routinely use to prolong and enhance life.

Now consider: How much money would someone have to pay you to give up for the rest of your life all the modern conveniences that Rockefeller lived without? Would you do it for $200 billion? Perhaps not. And if you wouldn’t, is it fair to say that you are better off than John D. Rockefeller, allegedly the richest American ever?

The preceding chapter discussed how standard price indexes, which are used to compare sums of money from different points in time, fail to fully reflect the introduction of new goods in the economy. As a result, the rate of inflation is overestimated. The flip side of this observation is that the rate of real economic growth is underestimated. Pondering Rockefeller’s life shows how significant this problem might be. Because of tremendous technological advances, the average American today is arguably “richer” than the richest American a century ago, even if that fact is lost in standard economic statistics.

PHOTO: © AP/WIDE WORLD PHOTOS

John D. Rockefeller

PRODUCTIVITY: ITS ROLE AND DETERMINANTS

Explaining the large variation in living standards around the world is, in one sense, very easy. As we will see, the explanation can be summarized in a single word—productivity. But in another sense, the international variation is deeply puzzling. To explain why incomes are so much higher in some countries than in others, we must look at the many factors that determine a nation’s productivity.

WHY PRODUCTIVITY IS SO IMPORTANT

Let’s begin our study of productivity and economic growth by developing a simple model based loosely on Daniel Defoe’s famous novel Robinson Crusoe about a sailor stranded on a desert island. Because Crusoe lives alone, he catches his own fish, grows his own vegetables, and makes his own clothes. We can think of Crusoe’s activities—his production and consumption of fish, vegetables, and
clothing—as a simple economy. By examining Crusoe’s economy, we can learn some lessons that also apply to more complex and realistic economies.

What determines Crusoe’s standard of living? In a word, productivity, the quantity of goods and services produced from each unit of labor input. If Crusoe is good at catching fish, growing vegetables, and making clothes, he lives well. If he is bad at doing these things, he lives poorly. Because Crusoe gets to consume only what he produces, his living standard is tied to his productivity.

In the case of Crusoe’s economy, it is easy to see that productivity is the key determinant of living standards and that growth in productivity is the key determinant of growth in living standards. The more fish Crusoe can catch per hour, the more he eats at dinner. If Crusoe finds a better place to catch fish, his productivity rises. This increase in productivity makes Crusoe better off: He can eat the extra fish, or he can spend less time fishing and devote more time to making other goods he enjoys.

Productivity’s key role in determining living standards is as true for nations as it is for stranded sailors. Recall that an economy’s gross domestic product (GDP) measures two things at once: the total income earned by everyone in the economy and the total expenditure on the economy’s output of goods and services. GDP can measure these two things simultaneously because, for the economy as a whole, they must be equal. Put simply, an economy’s income is the economy’s output.

Like Crusoe, a nation can enjoy a high standard of living only if it can produce a large quantity of goods and services. Americans live better than Nigerians because American workers are more productive than Nigerian workers. The Japanese have enjoyed more rapid growth in living standards than Argentineans because Japanese workers have experienced more rapidly growing productivity. Indeed, one of the Ten Principles of Economics in Chapter 1 is that a country’s standard of living depends on its ability to produce goods and services.

Hence, to understand the large differences in living standards we observe across countries or over time, we must focus on the production of goods and services. But seeing the link between living standards and productivity is only the first step. It leads naturally to the next question: Why are some economies so much better at producing goods and services than others?

**How Productivity Is Determined**

Although productivity is uniquely important in determining Robinson Crusoe’s standard of living, many factors determine Crusoe’s productivity. Crusoe will be better at catching fish, for instance, if he has more fishing poles, if he has been trained in the best fishing techniques, if his island has a plentiful fish supply, or if he invents a better fishing lure. Each of these determinants of Crusoe’s productivity—which we can call physical capital, human capital, natural resources, and technological knowledge—has a counterpart in more complex and realistic economies. Let’s consider each factor in turn.

**Physical Capital per Worker** Workers are more productive if they have tools with which to work. The stock of equipment and structures used to produce goods and services is called physical capital, or just capital. For example, when woodworkers make furniture, they use saws, lathes, and drill presses. More tools allow the woodworkers to produce their output more quickly and more accurately:
A worker with only basic hand tools can make less furniture each week than a worker with sophisticated and specialized woodworking equipment.

As you may recall, the inputs used to produce goods and services—labor, capital, and so on—are called the factors of production. An important feature of capital is that it is a produced factor of production. That is, capital is an input into the production process that in the past was an output from the production process. The woodworker uses a lathe to make the leg of a table. Earlier, the lathe itself was the output of a firm that manufactures lathes. The lathe manufacturer in turn used other equipment to make its product. Thus, capital is a factor of production used to produce all kinds of goods and services, including more capital.

**Human Capital per Worker**  A second determinant of productivity is human capital. Human capital is the economist’s term for the knowledge and skills that workers acquire through education, training, and experience. Human capital includes the skills accumulated in early childhood programs, grade school, high school, college, and on-the-job training for adults in the labor force.

Education, training, and experience are less tangible than lathes, bulldozers, and buildings, but human capital is like physical capital in many ways. Like physical capital, human capital raises a nation’s ability to produce goods and services. Also like physical capital, human capital is a produced factor of production. Producing human capital requires inputs in the form of teachers, libraries, and student time. Indeed, students can be viewed as “workers” who have the important job of producing the human capital that will be used in future production.

**Natural Resources per Worker**  A third determinant of productivity is natural resources. Natural resources are inputs into production that are provided by nature, such as land, rivers, and mineral deposits. Natural resources take two forms: renewable and nonrenewable. A forest is an example of a renewable resource. When one tree is cut down, a seedling can be planted in its place to be harvested in the future. Oil is an example of a nonrenewable resource. Because oil is produced by nature over many millions of years, there is only a limited supply. Once the supply of oil is depleted, it is impossible to create more.

Differences in natural resources are responsible for some of the differences in standards of living around the world. The historical success of the United States was driven in part by the large supply of land well suited for agriculture. Today, some countries in the Middle East, such as Kuwait and Saudi Arabia, are rich simply because they happen to be on top of some of the largest pools of oil in the world.

Although natural resources can be important, they are not necessary for an economy to be highly productive in producing goods and services. Japan, for instance, is one of the richest countries in the world, despite having few natural resources. International trade makes Japan’s success possible. Japan imports many of the natural resources it needs, such as oil, and exports its manufactured goods to economies rich in natural resources.

**Technological Knowledge**  A fourth determinant of productivity is technological knowledge—the understanding of the best ways to produce goods and services. A hundred years ago, most Americans worked on farms because farm technology required a high input of labor to feed the entire population. Today, thanks to advances in the technology of farming, a small fraction of the popu-
lation can produce enough food to feed the entire country. This technological change made labor available to produce other goods and services.

Technological knowledge takes many forms. Some technology is common knowledge—after one person uses it, everyone becomes aware of it. For example, once Henry Ford successfully introduced production in assembly lines, other carmakers quickly followed suit. Other technology is proprietary—it is known only by the company that discovers it. Only the Coca-Cola Company, for instance, knows the secret recipe for making its famous soft drink. Still other technology is proprietary for a short time. When a pharmaceutical company discovers a new drug, the patent system gives that company a temporary right to be its exclusive manufacturer. When the patent expires, however, other companies are allowed to make the drug. All these forms of technological knowledge are important for the economy’s production of goods and services.

It is worthwhile to distinguish between technological knowledge and human capital. Although they are closely related, there is an important difference. Technological knowledge refers to society’s understanding about how the world works. Human capital refers to the resources expended transmitting this understanding to the labor force. To use a relevant metaphor, knowledge is the quality of society’s

CHAPTER 17 PRODUCTION AND GROWTH 373

The Production Function

Economists often use a production function to describe the relationship between the quantity of inputs used in production and the quantity of output from production. For example, suppose $Y$ denotes the quantity of output, $L$ the quantity of labor, $K$ the quantity of physical capital, $H$ the quantity of human capital, and $N$ the quantity of natural resources. Then we might write

$$Y = A F(L, K, H, N),$$

where $F(\cdot)$ is a function that shows how the inputs are combined to produce output. $A$ is a variable that reflects the available production technology. As technology improves, $A$ rises, so the economy produces more output from any given combination of inputs.

Many production functions have a property called constant returns to scale. If a production function has constant returns to scale, then doubling all inputs causes the amount of output to double as well. Mathematically, we write that a production function has constant returns to scale if, for any positive number $x$,

$$xY = A F(xL, xK, xH, xN).$$

A doubling of all inputs would be represented in this equation by $x = 2$. The right side shows the inputs doubling, and the left side shows output doubling.

Production functions with constant returns to scale have an interesting and useful implication. To see this implication, it will prove instructive to set $x = 1/L$. Then the preceding equation becomes

$$Y/L = A F(1, K/L, H/L, N/L).$$

Notice that $Y/L$ is output per worker, which is a measure of productivity. This equation says that labor productivity depends on physical capital per worker ($K/L$), human capital per worker ($H/L$), and natural resources per worker ($N/L$). Productivity also depends on the state of technology, as reflected by the variable $A$. Thus, this equation provides a mathematical summary of the four determinants of productivity we have just discussed.
The concept of “capital” is sometimes interpreted broadly.

Measuring Capital

The Secrets of Intangible Wealth
By Ronald Bailey

A Mexican migrant to the U.S. is five times more productive than one who stays home. Why is that?

The answer is not the obvious one: This country has more machinery or tools or natural resources. Instead, according to some remarkable but largely ignored research—by the World Bank, of all places—it is because the average American has access to over $418,000 in intangible wealth, while the stay-at-home Mexican’s intangible wealth is just $34,000.

But what is intangible wealth, and how on earth is it measured? And what does it mean for the world’s people—poor and rich? That’s where the story gets even more interesting.

Two years ago the World Bank’s environmental economics department set out to assess the relative contributions of various kinds of capital to economic development. Its study, “Where Is the Wealth of Nations?: Measuring Capital for the 21st Century,” began by defining natural capital as the sum of nonrenewable resources (including oil, natural gas, coal and mineral resources), cropland, pasture land, forested areas and protected areas. Produced, or built, capital is what many of us think of when we think of capital: the sum of machinery, equipment, and structures (including infrastructure) and urban land.

But once the value[s] of all these are added up, the economists found something big was still missing: the vast majority of [the] world’s wealth! If one simply adds up the current value of a country’s natural resources and produced, or built, capital, there’s no way that can account for that country’s level of income.

The rest is the result of “intangible” factors—such as the trust among people in a society, an efficient judicial system, clear property rights and effective government. All this intangible capital also boosts the productivity of labor and results in higher total wealth. In fact, the World Bank finds, “Human capital and the value of institutions (as measured by rule of law) constitute the largest share of wealth in virtually all countries.”

Once one takes into account all of the world’s natural resources and produced capital, 80% of the wealth of rich countries...
and 60% of the wealth of poor countries is of this intangible type. The bottom line: “Rich countries are largely rich because of the skills of their populations and the quality of the institutions supporting economic activity.”

What the World Bank economists have brilliantly done is quantify the intangible value of education and social institutions. According to their regression analyses, for example, the rule of law explains 57% of countries’ intangible capital. Education accounts for 36%.

The rule-of-law index was devised using several hundred individual variables measuring perceptions of governance, drawn from 25 separate data sources constructed by 18 different organizations. The latter include civil society groups (Freedom House), political and business risk-rating agencies (Economist Intelligence Unit) and think tanks (International Budget Project Open Budget Index).

Switzerland scores 99.5 out of 100 on the rule-of-law index and the U.S. hits 91.8. By contrast, Nigeria’s score is a pitiful 5.8; Burundi’s 4.3; and Ethiopia’s 16.4. The members of the Organization for Economic Cooperation and Development—30 wealthy developed countries—have an average score of 90, while sub-Saharan Africa’s is a dismal 28.

The natural wealth in rich countries like the U.S. is a tiny proportion of their overall wealth—typically 1% to 3%—yet they derive more value from what they have. Cropland, pastures and forests are more valuable in rich countries because they can be combined with other capital like machinery and strong property rights to produce more value. Machinery, buildings, roads and so forth account for 17% of the rich countries’ total wealth.

Overall, the average per capita wealth in the rich Organization for Economic Cooperation Development (OECD) countries is $440,000, consisting of $10,000 in natural capital, $76,000 in produced capital, and a whopping $354,000 in intangible capital. (Switzerland has the highest per capita wealth, at $648,000. The U.S. is fourth at $513,000.)

By comparison, the World Bank study finds that total wealth for the low income countries averages $7,216 per person. That consists of $2,075 in natural capital, $1,150 in produced capital and $3,991 in intangible capital. The countries with the lowest per capita wealth are Ethiopia ($1,965), Nigeria ($2,748), and Burundi ($2,859).

In fact, some countries are so badly run, that they actually have negative intangible capital. Through rampant corruption and failing school systems, Nigeria and the Democratic Republic of the Congo are destroying their intangible capital and ensuring that their people will be poorer in the future.

The World Bank’s pathbreaking “Where Is the Wealth of Nations?” convincingly demonstrates that the “mainsprings of development” are the rule of law and a good school system. The big question that its researchers don’t answer is: How can the people of the developing world rid themselves of the kleptocrats who loot their countries and keep them poor?

people advocated mandatory recycling and rationing of tin and copper so that supplies would be available for future generations. Today, however, plastic has replaced tin as a material for making many food containers, and phone calls often travel over fiber-optic cables, which are made from sand. Technological progress has made once crucial natural resources less necessary.

But are all these efforts enough to permit continued economic growth? One way to answer this question is to look at the prices of natural resources. In a market economy, scarcity is reflected in market prices. If the world were running out of natural resources, then the prices of those resources would be rising over time. But in fact, the opposite is more often true. Natural resource prices exhibit substantial short-run fluctuations, but over long spans of time, the prices of most natural resources (adjusted for overall inflation) are stable or falling. It appears that our ability to conserve these resources is growing more rapidly than their supplies are dwindling. Market prices give no reason to believe that natural resources are a limit to economic growth.

**Quick Quiz** List and describe four determinants of a country’s productivity.

---

**ECONOMIC GROWTH AND PUBLIC POLICY**

So far, we have determined that a society’s standard of living depends on its ability to produce goods and services and that its productivity in turn depends on physical capital per worker, human capital per worker, natural resources per worker, and technological knowledge. Let’s now turn to the question faced by policymakers around the world: What can government policy do to raise productivity and living standards?

**Saving and Investment**

Because capital is a produced factor of production, a society can change the amount of capital it has. If today the economy produces a large quantity of new capital goods, then tomorrow it will have a larger stock of capital and be able to produce more goods and services. Thus, one way to raise future productivity is to invest more current resources in the production of capital.

One of the *Ten Principles of Economics* presented in Chapter 1 is that people face trade-offs. This principle is especially important when considering the accumulation of capital. Because resources are scarce, devoting more resources to producing capital requires devoting fewer resources to producing goods and services for current consumption. That is, for society to invest more in capital, it must consume less and save more of its current income. The growth that arises from capital accumulation is not a free lunch: It requires that society sacrifice consumption of goods and services in the present to enjoy higher consumption in the future.

The next chapter examines in more detail how the economy’s financial markets coordinate saving and investment. It also examines how government policies influence the amount of saving and investment that takes place. At this point, it is important to note that encouraging saving and investment is one way that a government can encourage growth and, in the long run, raise the economy’s standard of living.
DIMINISHING RETURNS AND THE CATCH-UP EFFECT

Suppose that a government pursues policies that raise the nation’s saving rate—the percentage of GDP devoted to saving rather than consumption. What happens? With the nation saving more, fewer resources are needed to make consumption goods, and more resources are available to make capital goods. As a result, the capital stock increases, leading to rising productivity and more rapid growth in GDP. But how long does this higher rate of growth last? Assuming that the saving rate remains at its new higher level, does the growth rate of GDP stay high indefinitely or only for a period of time?

The traditional view of the production process is that capital is subject to diminishing returns: As the stock of capital rises, the extra output produced from an additional unit of capital falls. In other words, when workers already have a large quantity of capital to use in producing goods and services, giving them an additional unit of capital increases their productivity only slightly. This is illustrated in Figure 1, which shows how the amount of capital per worker determines the amount of output per worker, holding constant all the other determinants of output.

Because of diminishing returns, an increase in the saving rate leads to higher growth only for a while. As the higher saving rate allows more capital to be accumulated, the benefits from additional capital become smaller over time, and so growth slows down. In the long run, the higher saving rate leads to a higher level of productivity and income but not to higher growth in these variables. Reaching this long run, however, can take quite a while. According to studies of international data

**Illustrating the Production Function**

This figure shows how the amount of capital per worker influences the amount of output per worker. Other determinants of output, including human capital, natural resources, and technology, are held constant. The curve becomes flatter as the amount of capital increases because of diminishing returns to capital.
on economic growth, increasing the saving rate can lead to substantially higher growth for a period of several decades.

The diminishing returns to capital has another important implication: Other things equal, it is easier for a country to grow fast if it starts out relatively poor. This effect of initial conditions on subsequent growth is sometimes called the **catch-up effect**. In poor countries, workers lack even the most rudimentary tools and, as a result, have low productivity. Small amounts of capital investment would substantially raise these workers’ productivity. By contrast, workers in rich countries have large amounts of capital with which to work, and this partly explains their high productivity. Yet with the amount of capital per worker already so high, additional capital investment has a relatively small effect on productivity. Studies of international data on economic growth confirm this catch-up effect: Controlling for other variables, such as the percentage of GDP devoted to investment, poor countries tend to grow at faster rates than rich countries.

This catch-up effect can help explain some otherwise puzzling facts. Here’s an example: From 1960 to 1990, the United States and South Korea devoted a similar share of GDP to investment. Yet over this time, the United States experienced only mediocre growth of about 2 percent, while South Korea experienced spectacular growth of more than 6 percent. The explanation is the catch-up effect. In 1960, South Korea had GDP per person less than one-tenth the U.S. level, in part because previous investment had been so low. With a small initial capital stock, the benefits to capital accumulation were much greater in South Korea, and this gave South Korea a higher subsequent growth rate.

This catch-up effect shows up in other aspects of life. When a school gives an end-of-year award to the “Most Improved” student, that student is usually one who began the year with relatively poor performance. Students who began the year not studying find improvement easier than students who always worked hard. Note that it is good to be “Most Improved,” given the starting point, but it is even better to be “Best Student.” Similarly, economic growth over the last several decades has been much more rapid in South Korea than in the United States, but GDP per person is still higher in the United States.

**INVESTMENT FROM ABROAD**

So far, we have discussed how policies aimed at increasing a country’s saving rate can increase investment and, thereby, long-term economic growth. Yet saving by domestic residents is not the only way for a country to invest in new capital. The other way is investment by foreigners.

Investment from abroad takes several forms. Ford Motor Company might build a car factory in Mexico. A capital investment that is owned and operated by a foreign entity is called **foreign direct investment**. Alternatively, an American might buy stock in a Mexican corporation (that is, buy a share in the ownership of the corporation); the Mexican corporation can use the proceeds from the stock sale to build a new factory. An investment that is financed with foreign money but operated by domestic residents is called **foreign portfolio investment**. In both cases, Americans provide the resources necessary to increase the stock of capital in Mexico. That is, American saving is being used to finance Mexican investment.

When foreigners invest in a country, they do so because they expect to earn a return on their investment. Ford’s car factory increases the Mexican capital stock and, therefore, increases Mexican productivity and Mexican GDP. Yet Ford takes
Investment from abroad, therefore, does not have the same effect on all measures of economic prosperity. Recall that gross domestic product (GDP) is the income earned within a country by both residents and nonresidents, whereas gross national product (GNP) is the income earned by residents of a country both at home and abroad. When Ford opens its car factory in Mexico, some of the income the factory generates accrues to people who do not live in Mexico. As a result, foreign investment in Mexico raises the income of Mexicans (measured by GNP) by less than it raises the production in Mexico (measured by GDP).

Nonetheless, investment from abroad is one way for a country to grow. Even though some of the benefits from this investment flow back to the foreign owners, this investment does increase the economy’s stock of capital, leading to higher productivity and higher wages. Moreover, investment from abroad is one way for poor countries to learn the state-of-the-art technologies developed and used in richer countries. For these reasons, many economists who advise governments in less developed economies advocate policies that encourage investment from abroad. Often, this means removing restrictions that governments have imposed on foreign ownership of domestic capital.

An organization that tries to encourage the flow of capital to poor countries is the World Bank. This international organization obtains funds from the world’s advanced countries, such as the United States, and uses these resources to make loans to less developed countries so that they can invest in roads, sewer systems, schools, and other types of capital. It also offers the countries advice about how the funds might best be used. The World Bank, together with its sister organization, the International Monetary Fund, was set up after World War II. One lesson from the war was that economic distress often leads to political turmoil, international tensions, and military conflict. Thus, every country has an interest in promoting economic prosperity around the world. The World Bank and the International Monetary Fund were established to achieve that common goal.

**Education**

Education—investment in human capital—is at least as important as investment in physical capital for a country’s long-run economic success. In the United States, each year of schooling has historically raised a person’s wage by an average of about 10 percent. In less developed countries, where human capital is especially scarce, the gap between the wages of educated and uneducated workers is even larger. Thus, one way government policy can enhance the standard of living is to provide good schools and to encourage the population to take advantage of them.

Investment in human capital, like investment in physical capital, has an opportunity cost. When students are in school, they forgo the wages they could have earned. In less developed countries, children often drop out of school at an early age, even though the benefit of additional schooling is very high, simply because their labor is needed to help support the family.

Some economists have argued that human capital is particularly important for economic growth because human capital conveys positive externalities. An *externality* is the effect of one person’s actions on the well-being of a bystander.
Promoting Human Capital

Human capital is a key to economic growth. With this in mind, some developing countries now give parents an immediate financial incentive to keep their children in school.

Brazil Pays Parents to Help Poor Be Pupils, Not Wage Earners

By Celia W. Dugger

ORTALEZA, Brazil—Vandelson Andrade, 13, often used to skip school to work 12-hour days on the small, graceful fishing boats that sail from the picturesque harbor here. His meager earnings helped pay for rice and beans for his desperately poor family.

But this year he qualified for a small monthly cash payment from the government that his mother receives on the condition that he shows up in the classroom.

“I can’t skip school anymore,” said Vandelson, whose hand-me-down pants were so big that the crotch ended at his knees and the legs bunched up around his ankles. “If I miss one more day, my mother won’t get the money.”

This year, Vandelson will finally pass the fourth grade on his third try—a small victory in a new breed of social program that is spreading swiftly across Latin America. It is a developing-country version of American welfare reform: to break the cycle of poverty, the government gives the poor small cash payments in exchange for keeping their children in school and taking them for regular medical checkups.

“I think these programs are as close as you can come to a magic bullet in development,” said Nancy Birdsall, president of the Center for Global Development, a nonprofit research group in Washington. “They’re creating an incentive for families to invest in their own children’s futures. Every decade or so, we see something that can really make a difference, and this is one of those things.”

Antônio Souza, 48, and Maria Torres, 37, are raising seven children in a mud hut a couple of hills away from Ms. Andrade. Every member of the family is sinewy and lean. The parents cannot remember the last time the family ate meat or vegetables. But their grant of $27 a month makes it possible to buy rice, sugar, pasta and oil.

Mr. Souza and Ms. Torres, illiterate believers in the power of education, have always sent their children to school. “If they don’t study, they’ll turn into dummies like me,” said their father, whose weathered, deeply creased face broke into a wide smile as he surveyed his bright-eyed daughters, Ana Paula, 11, and Daniele, 8, among them. “All I can do is work in the fields.”

His wife said proudly: “There are fathers who don’t want their children to go to school. But this man here has done everything he could to send his children to school.”

An educated person, for instance, might generate new ideas about how best to produce goods and services. If these ideas enter society’s pool of knowledge so everyone can use them, then the ideas are an external benefit of education. In this case, the return to schooling for society is even greater than the return for the individual. This argument would justify the large subsidies to human-capital investment that we observe in the form of public education.
One problem facing some poor countries is the brain drain—the emigration of many of the most highly educated workers to rich countries, where these workers can enjoy a higher standard of living. If human capital does have positive externalities, then this brain drain makes those people left behind poorer than they otherwise would be. This problem offers policymakers a dilemma. On the one hand, the United States and other rich countries have the best systems of higher education, and it would seem natural for poor countries to send their best students abroad to earn higher degrees. On the other hand, those students who have spent time abroad may choose not to return home, and this brain drain will reduce the poor nation’s stock of human capital even further.

**Health and Nutrition**

The term human capital usually refers to education, but it can also be used to describe another type of investment in people: expenditures that lead to a healthier population. Other things equal, healthier workers are more productive. The right investments in the health of the population provide one way for a nation to increase productivity and raise living standards.

Economic historian Robert Fogel has suggested that a significant factor in long-run economic growth is improved health from better nutrition. He estimates that in Great Britain in 1780, about one in five people were so malnourished that they were incapable of manual labor. Among those who could work, insufficient caloric intake substantially reduced the work effort they could put forth. As nutrition improved, so did workers’ productivity.

Fogel studies these historical trends in part by looking at the height of the population. Short stature can be an indicator of malnutrition, especially during gestation and the early years of life. Fogel finds that as nations develop economically, people eat more, and the population gets taller. From 1775 to 1975, the average caloric intake in Great Britain rose by 26 percent, and the height of the average man rose by 3.6 inches. Similarly, during the spectacular economic growth in South Korea from 1962 to 1995, caloric consumption rose by 44 percent, and average male height rose by 2 inches. Of course, a person’s height is determined by a combination of genetic predisposition and environment. But because the genetic makeup of a population is slow to change, such increases in average height are most likely due to changes in the environment—nutrition being the obvious explanation.

Moreover, studies have found that height is an indicator of productivity. Looking at data on a large number of workers at a point in time, researchers have found that taller workers tend to earn more. Because wages reflect a worker’s productivity, this finding suggests that taller workers tend to be more productive. The effect of height on wages is especially pronounced in poorer countries, where malnutrition is a bigger risk.

Fogel won the Nobel Prize in Economics in 1993 for his work in economic history, which includes not only his studies of nutrition but also his studies of American slavery and the role of railroads in the development of the American economy. In the lecture he gave when he was awarded the prize, he surveyed the evidence on health and economic growth. He concluded that “improved gross nutrition accounts for roughly 30 percent of the growth of per capita income in Britain between 1790 and 1980.”
Today, malnutrition is fortunately rare in developed nations such as Great Britain and the United States. (Obesity is a more widespread problem.) But for people in developing nations, poor health and inadequate nutrition remain obstacles to higher productivity and improved living standards. The United Nations estimates that almost a third of the population in sub-Saharan Africa is undernourished.

The causal link between health and wealth runs in both directions. Poor countries are poor in part because their populations are not healthy, and their populations are not healthy in part because they are poor and cannot afford adequate healthcare and nutrition. It is a vicious circle. But this fact opens the possibility of a virtuous circle: Policies that lead to more rapid economic growth would naturally improve health outcomes, which in turn would further promote economic growth.

**Property Rights and Political Stability**

Another way policymakers can foster economic growth is by protecting property rights and promoting political stability. This issue goes to the very heart of how market economies work.

Production in market economies arises from the interactions of millions of individuals and firms. When you buy a car, for instance, you are buying the output of a car dealer, a car manufacturer, a steel company, an iron ore mining company, and so on. This division of production among many firms allows the economy’s factors of production to be used as effectively as possible. To achieve this outcome, the economy has to coordinate transactions among these firms, as well as between firms and consumers. Market economies achieve this coordination through market prices. That is, market prices are the instrument with which the invisible hand of the marketplace brings supply and demand into balance in each of the many thousands of markets that make up the economy.

An important prerequisite for the price system to work is an economy-wide respect for *property rights*. Property rights refer to the ability of people to exercise authority over the resources they own. A mining company will not make the effort to mine iron ore if it expects the ore to be stolen. The company mines the ore only if it is confident that it will benefit from the ore’s subsequent sale. For this reason, courts serve an important role in a market economy: They enforce property rights. Through the criminal justice system, the courts discourage direct theft. In addition, through the civil justice system, the courts ensure that buyers and sellers live up to their contracts.

Those of us in developed countries tend to take property rights for granted, but those living in less developed countries understand that a lack of property rights can be a major problem. In many countries, the system of justice does not work well. Contracts are hard to enforce, and fraud often goes unpunished. In more extreme cases, the government not only fails to enforce property rights but actually infringes upon them. To do business in some countries, firms are expected to bribe powerful government officials. Such corruption impedes the coordinating power of markets. It also discourages domestic saving and investment from abroad.

One threat to property rights is political instability. When revolutions and coups are common, there is doubt about whether property rights will be respected in the future. If a revolutionary government might confiscate the capital of some businesses, as was often true after communist revolutions, domestic residents have
less incentive to save, invest, and start new businesses. At the same time, foreigners have less incentive to invest in the country. Even the threat of revolution can act to depress a nation’s standard of living.

Thus, economic prosperity depends in part on political prosperity. A country with an efficient court system, honest government officials, and a stable constitution will enjoy a higher economic standard of living than a country with a poor court system, corrupt officials, and frequent revolutions and coups.

**Free Trade**

Some of the world’s poorest countries have tried to achieve more rapid economic growth by pursuing *inward-oriented policies*. These policies attempt to increase productivity and living standards within the country by avoiding interaction with the rest of the world. Domestic firms often advance the infant-industry argument, claiming they need protection from foreign competition to thrive and grow. Together with a general distrust of foreigners, this argument has at times led policymakers in less developed countries to impose tariffs and other trade restrictions.

Most economists today believe that poor countries are better off pursuing *outward-oriented policies* that integrate these countries into the world economy. International trade in goods and services can improve the economic well-being of a country’s citizens. Trade is, in some ways, a type of technology. When a country exports wheat and imports textiles, the country benefits as if it had invented a technology for turning wheat into textiles. A country that eliminates trade restrictions will, therefore, experience the same kind of economic growth that would occur after a major technological advance.

The adverse impact of inward orientation becomes clear when one considers the small size of many less developed economies. The total GDP of Argentina, for instance, is about that of Philadelphia. Imagine what would happen if the Philadelphia city council were to prohibit city residents from trading with people living outside the city limits. Without being able to take advantage of the gains from trade, Philadelphia would need to produce all the goods it consumes. It would also have to produce all its own capital goods, rather than importing state-of-the-art equipment from other cities. Living standards in Philadelphia would fall immediately, and the problem would likely only get worse over time. This is precisely what happened when Argentina pursued inward-oriented policies throughout much of the 20th century. By contrast, countries that pursued outward-oriented policies, such as South Korea, Singapore, and Taiwan, enjoyed high rates of economic growth.

The amount that a nation trades with others is determined not only by government policy but also by geography. Countries with natural seaports find trade easier than countries without this resource. It is not a coincidence that many of the world’s major cities, such as New York, San Francisco, and Hong Kong, are located next to oceans. Similarly, because landlocked countries find international trade more difficult, they tend to have lower levels of income than countries with easy access to the world’s waterways. For example, countries with more than 80 percent of their population living within 100 kilometers of a coast have an average GDP per person about four times as large as countries with less than 20 percent of their population living near a coast. The critical importance of access to the sea helps explain why the African continent, which contains many landlocked countries, is so poor.
RESEARCH AND DEVELOPMENT

The primary reason that living standards are higher today than they were a century ago is that technological knowledge has advanced. The telephone, the transistor, the computer, and the internal combustion engine are among the thousands of innovations that have improved the ability to produce goods and services.

Although most technological advances come from private research by firms and individual inventors, there is also a public interest in promoting these efforts. To a large extent, knowledge is a public good: Once one person discovers an idea, the idea enters society’s pool of knowledge, and other people can freely use it. Just as government has a role in providing a public good such as national defense, it also has a role in encouraging the research and development of new technologies.

The U.S. government has long played a role in the creation and dissemination of technological knowledge. A century ago, the government sponsored research about farming methods and advised farmers how best to use their land. More recently, the U.S. government, through the Air Force and NASA, has supported aerospace research; as a result, the United States is a leading maker of rockets and planes. The government continues to encourage advances in knowledge with research grants from the National Science Foundation and the National Institutes of Health and with tax breaks for firms engaging in research and development.

Yet another way in which government policy encourages research is through the patent system. When a person or firm invents a new product, such as a new drug, the inventor can apply for a patent. If the product is deemed truly original, the government awards the patent, which gives the inventor the exclusive right to make the product for a specified number of years. In essence, the patent gives the inventor a property right over his invention, turning his new idea from a public good into a private good. By allowing inventors to profit from their inventions—even if only temporarily—the patent system enhances the incentive for individuals and firms to engage in research.

POPULATION GROWTH

Economists and other social scientists have long debated how population affects a society. The most direct effect is on the size of the labor force: A large population means more workers to produce goods and services. The tremendous size of the Chinese population is one reason China is such an important player in the world economy.

At the same time, however, a large population means more people to consume those goods and services. So while a large population means a larger total output of goods and services, it need not mean a higher standard of living for a typical citizen. Indeed, both large and small nations are found at all levels of economic development.

Beyond these obvious effects of population size, population growth interacts with the other factors of production in ways that are more subtle and open to debate.

Stretching Natural Resources  Thomas Robert Malthus (1766–1834), an English minister and early economic thinker, is famous for his book called An Essay on the Principle of Population as It Affects the Future Improvement of Society. In it,
he offered what may be history’s most chilling forecast. Malthus argued that an ever-increasing population would continually strain society’s ability to provide for itself. As a result, mankind was doomed to forever live in poverty.

Malthus’s logic was simple. He began by noting that “food is necessary to the existence of man” and that “the passion between the sexes is necessary and will remain nearly in its present state.” He concluded that “the power of population is infinitely greater than the power in the earth to produce subsistence for man.” According to Malthus, the only check on population growth was “misery and vice.” Attempts by charities or governments to alleviate poverty were counterproductive, he argued, because they merely allowed the poor to have more children, placing even greater strains on society’s productive capabilities.

Malthus may have correctly described the world at the time when he lived, but fortunately, his dire forecast was far off the mark. The world population has increased about sixfold over the past two centuries, but living standards around the world are on average much higher. As a result of economic growth, chronic hunger and malnutrition are less common now than they were in Malthus’s day. When famines occur from time to time, they are more often the result of an unequal income distribution or political instability than inadequate food production.

Where did Malthus go wrong? As we discussed in a case study earlier in this chapter, growth in human ingenuity has offset the effects of a larger population. Pesticides, fertilizers, mechanized farm equipment, new crop varieties, and other technological advances that Malthus never imagined have allowed each farmer to feed ever greater numbers of people. Even with more mouths to feed, fewer farmers are necessary because each farmer is much more productive.

**Diluting the Capital Stock**  Whereas Malthus worried about the effects of population on the use of natural resources, some modern theories of economic growth emphasize its effects on capital accumulation. According to these theories, high population growth reduces GDP per worker because rapid growth in the number of workers forces the capital stock to be spread more thinly. In other words, when population growth is rapid, each worker is equipped with less capital. A smaller quantity of capital per worker leads to lower productivity and lower GDP per worker.

This problem is most apparent in the case of human capital. Countries with high population growth have large numbers of school-age children. This places a larger burden on the educational system. It is not surprising, therefore, that educational attainment tends to be low in countries with high population growth.

The differences in population growth around the world are large. In developed countries, such as the United States and those in Western Europe, the population has risen only about 1 percent per year in recent decades and is expected to rise even more slowly in the future. By contrast, in many poor African countries, population grows at about 3 percent per year. At this rate, the population doubles every 23 years. This rapid population growth makes it harder to provide workers with the tools and skills they need to achieve high levels of productivity.

Although rapid population growth is not the main reason that less developed countries are poor, some analysts believe that reducing the rate of population growth would help these countries raise their standards of living. In some countries, this goal is accomplished directly with laws that regulate the number of children families may have. China, for instance, allows only one child per family;
couples who violate this rule are subject to substantial fines. In countries with greater freedom, the goal of reduced population growth is accomplished less directly by increasing awareness of birth control techniques.

Another way in which a country can influence population growth is to apply one of the Ten Principles of Economics: People respond to incentives. Bearing a child, like any decision, has an opportunity cost. When the opportunity cost rises, people will choose to have smaller families. In particular, women with the opportunity to receive a good education and desirable employment tend to want fewer children than those with fewer opportunities outside the home. Hence, policies that foster equal treatment of women may be one way for less developed economies to reduce the rate of population growth and, perhaps, raise their standards of living.
Promoting Technological Progress  Although rapid population growth may depress economic prosperity by reducing the amount of capital each worker has, it may also have some benefits. Some economists have suggested that world population growth has been an engine of technological progress and economic prosperity. The mechanism is simple: If there are more people, then there are more scientists, inventors, and engineers to contribute to technological advance, which benefits everyone.

Economist Michael Kremer has provided some support for this hypothesis in an article titled “Population Growth and Technological Change: One Million B.C. to 1990,” which was published in the *Quarterly Journal of Economics* in 1993. Kremer begins by noting that over the broad span of human history, world growth rates have increased with world population. For example, world growth was more
rapid when the world population was 1 billion (which occurred around the year 1800) than when the population was only 100 million (around 500 B.C.). This fact is consistent with the hypothesis that a larger population induces more technological progress.

Kremer’s second piece of evidence comes from comparing regions of the world. The melting of the polar icecaps at the end of the Ice Age around 10,000 B.C. flooded the land bridges and separated the world into several distinct regions that could not communicate with one another for thousands of years. If technological progress is more rapid when there are more people to discover things, then larger regions should have experienced more rapid growth.

According to Kremer, that is exactly what happened. The most successful region of the world in 1500 (when Columbus reestablished technological contact) comprised the “Old World” civilizations of the large Eurasia-Africa region. Next in technological development were the Aztec and Mayan civilizations in the Americas, followed by the hunter-gatherers of Australia, and then the primitive people of Tasmania, who lacked even fire-making and most stone and bone tools.

The smallest isolated region was Flinders Island, a tiny island between Tasmania and Australia. With the smallest population, Flinders Island had the fewest opportunities for technological advance and, indeed, seemed to regress. Around 3000 B.C., human society on Flinders Island died out completely. A large population, Kremer concludes, is a prerequisite for technological advance.

**Quick Quiz** Describe three ways a government policymaker can try to raise the growth in living standards in a society. Are there any drawbacks to these policies?

---

**CONCLUSION: THE IMPORTANCE OF LONG-RUN GROWTH**

In this chapter, we have discussed what determines the standard of living in a nation and how policymakers can endeavor to raise the standard of living through policies that promote economic growth. Most of this chapter is summarized in one of the Ten Principles of Economics: A country’s standard of living depends on its ability to produce goods and services. Policymakers who want to encourage growth in living standards must aim to increase their nation’s productive ability by encouraging rapid accumulation of the factors of production and ensuring that these factors are employed as effectively as possible.

Economists differ in their views of the role of government in promoting economic growth. At the very least, government can lend support to the invisible hand by maintaining property rights and political stability. More controversial is whether government should target and subsidize specific industries that might be especially important for technological progress. There is no doubt that these issues are among the most important in economics. The success of one generation’s policymakers in learning and heeding the fundamental lessons about economic growth determines what kind of world the next generation will inherit.
SUMMARY

- Economic prosperity, as measured by GDP per person, varies substantially around the world. The average income in the world’s richest countries is more than ten times that in the world’s poorest countries. Because growth rates of real GDP also vary substantially, the relative positions of countries can change dramatically over time.

- The standard of living in an economy depends on the economy’s ability to produce goods and services. Productivity, in turn, depends on the physical capital, human capital, natural resources, and technological knowledge available to workers.

- Government policies can try to influence the economy’s growth rate in many ways: by encouraging saving and investment, encouraging investment from abroad, fostering education, promoting good health, maintaining property rights and political stability, allowing free trade, and promoting the research and development of new technologies.

- The accumulation of capital is subject to diminishing returns: The more capital an economy has, the less additional output the economy gets from an extra unit of capital. As a result, while higher saving leads to higher growth for a period of time, growth eventually slows down as capital, productivity, and income rise. Also because of diminishing returns, the return to capital is especially high in poor countries. Other things equal, these countries can grow faster because of the catch-up effect.

- Population growth has a variety of effects on economic growth. On the one hand, more rapid population growth may lower productivity by stretching the supply of natural resources and by reducing the amount of capital available for each worker. On the other hand, a larger population may enhance the rate of technological progress because there are more scientists and engineers.

KEY CONCEPTS

productivity, p. 371  
physical capital, p. 371  
human capital, p. 372  
natural resources, p. 372  
technological knowledge, p. 372  
diminishing returns, p. 377  
catch-up effect, p. 378

QUESTIONS FOR REVIEW

1. What does the level of a nation’s GDP measure? What does the growth rate of GDP measure? Would you rather live in a nation with a high level of GDP and a low growth rate or in a nation with a low level of GDP and a high growth rate?

2. List and describe four determinants of productivity.

3. In what way is a college degree a form of capital?

4. Explain how higher saving leads to a higher standard of living. What might deter a policymaker from trying to raise the rate of saving?

5. Does a higher rate of saving lead to higher growth temporarily or indefinitely?

6. Why would removing a trade restriction, such as a tariff, lead to more rapid economic growth?

7. How does the rate of population growth influence the level of GDP per person?

8. Describe two ways the U.S. government tries to encourage advances in technological knowledge.
PROBLEMS AND APPLICATIONS

1. Most countries, including the United States, import substantial amounts of goods and services from other countries. Yet the chapter says that a nation can enjoy a high standard of living only if it can produce a large quantity of goods and services itself. Can you reconcile these two facts?

2. Suppose that society decided to reduce consumption and increase investment.
   a. How would this change affect economic growth?
   b. What groups in society would benefit from this change? What groups might be hurt?

3. Societies choose what share of their resources to devote to consumption and what share to devote to investment. Some of these decisions involve private spending; others involve government spending.
   a. Describe some forms of private spending that represent consumption and some forms that represent investment. The national income accounts include tuition as a part of consumer spending. In your opinion, are the resources you devote to your education a form of consumption or a form of investment?
   b. Describe some forms of government spending that represent consumption and some forms that represent investment. In your opinion, should we view government spending on health programs as a form of consumption or investment? Would you distinguish between health programs for the young and health programs for the elderly?

4. What is the opportunity cost of investing in capital? Do you think a country can “overinvest” in capital? What is the opportunity cost of investing in human capital? Do you think a country can “overinvest” in human capital? Explain.

5. Suppose that an auto company owned entirely by German citizens opens a new factory in South Carolina.
   a. What sort of foreign investment would this represent?
   b. What would be the effect of this investment on U.S. GDP? Would the effect on U.S. GNP be larger or smaller?

6. In the 1990s and the first decade of the 2000s, investors from the Asian economies of Japan and China made significant direct and portfolio investments in the United States. At the time, many Americans were unhappy that this investment was occurring.
   a. In what way was it better for the United States to receive this foreign investment than not to receive it?
   b. In what way would it have been better still for Americans to have made this investment?

7. In many developing nations, young women have lower enrollment rates in secondary school than do young men. Describe several ways in which greater educational opportunities for young women could lead to faster economic growth in these countries.

8. International data show a positive correlation between income per person and the health of the population.
   a. Explain how higher income might cause better health outcomes.
   b. Explain how better health outcomes might cause higher income.
   c. How might the relative importance of your two hypotheses be relevant for public policy?

9. International data show a positive correlation between political stability and economic growth.
   a. Through what mechanism could political stability lead to strong economic growth?
   b. Through what mechanism could strong economic growth lead to political stability?

10. From 1950 to 2000, manufacturing employment as a percentage of total employment in the U.S. economy fell from 28 percent to 13 percent. At the same time, manufacturing output experienced slightly more rapid growth than the overall economy.
    a. What do these facts say about growth in labor productivity (defined as output per worker) in manufacturing?
    b. In your opinion, should policymakers be concerned about the decline in the share of manufacturing employment? Explain.
Saving, Investment, and the Financial System

Imagine that you have just graduated from college (with a degree in economics, of course) and you decide to start your own business—an economic forecasting firm. Before you make any money selling your forecasts, you have to incur substantial costs to set up your business. You have to buy computers with which to make your forecasts, as well as desks, chairs, and filing cabinets to furnish your new office. Each of these items is a type of capital that your firm will use to produce and sell its services.

How do you obtain the funds to invest in these capital goods? Perhaps you are able to pay for them out of your past savings. More likely, however, like most entrepreneurs, you do not have enough money of your own to finance the start of your business. As a result, you have to get the money you need from other sources.

There are various ways to finance these capital investments. You could borrow the money, perhaps from a bank or from a friend or relative. In this case, you would promise not only to return the money at a later date but also to pay interest for the use of the money. Alternatively, you could convince someone to provide the money you need for your business in exchange for a share of your future profits, whatever they might happen to be. In either case, your investment in computers and office equipment is being financed by someone else’s saving.
The financial system consists of those institutions that help to match one person’s saving with another person’s investment. As we discussed in the previous chapter, saving and investment are key ingredients to long-run economic growth: When a country saves a large portion of its GDP, more resources are available for investment in capital, and higher capital raises a country’s productivity and living standard. The previous chapter, however, did not explain how the economy coordinates saving and investment. At any time, some people want to save some of their income for the future, and others want to borrow to finance investments in new and growing businesses. What brings these two groups of people together? What ensures that the supply of funds from those who want to save balances the demand for funds from those who want to invest?

This chapter examines how the financial system works. First, we discuss the large variety of institutions that make up the financial system in our economy. Second, we discuss the relationship between the financial system and some key macroeconomic variables—notably saving and investment. Third, we develop a model of the supply and demand for funds in financial markets. In the model, the interest rate is the price that adjusts to balance supply and demand. The model shows how various government policies affect the interest rate and, thereby, society’s allocation of scarce resources.

FINANCIAL INSTITUTIONS IN THE U.S. ECONOMY

At the broadest level, the financial system moves the economy’s scarce resources from savers (people who spend less than they earn) to borrowers (people who spend more than they earn). Savers save for various reasons—to put a child through college in several years or to retire comfortably in several decades. Similarly, borrowers borrow for various reasons—to buy a house in which to live or to start a business with which to make a living. Savers supply their money to the financial system with the expectation that they will get it back with interest at a later date. Borrowers demand money from the financial system with the knowledge that they will be required to pay it back with interest at a later date.

The financial system is made up of various financial institutions that help coordinate savers and borrowers. As a prelude to analyzing the economic forces that drive the financial system, let’s discuss the most important of these institutions. Financial institutions can be grouped into two categories: financial markets and financial intermediaries. We consider each category in turn.

FINANCIAL MARKETS

Financial markets are the institutions through which a person who wants to save can directly supply funds to a person who wants to borrow. The two most important financial markets in our economy are the bond market and the stock market.

The Bond Market When Intel, the giant maker of computer chips, wants to borrow to finance construction of a new factory, it can borrow directly from the public. It does this by selling bonds. A bond is a certificate of indebtedness that specifies the obligations of the borrower to the holder of the bond. Put simply, a bond is an IOU. It identifies the time at which the loan will be repaid, called the
date of maturity, and the rate of interest that will be paid periodically until the loan matures. The buyer of a bond gives his or her money to Intel in exchange for this promise of interest and eventual repayment of the amount borrowed (called the principal). The buyer can hold the bond until maturity or can sell the bond at an earlier date to someone else.

There are literally millions of different bonds in the U.S. economy. When large corporations, the federal government, or state and local governments need to borrow to finance the purchase of a new factory, a new jet fighter, or a new school, they usually do so by issuing bonds. If you look at The Wall Street Journal or the business section of your local newspaper, you will find a listing of the prices and interest rates on some of the most important bond issues. These bonds differ according to three significant characteristics.

The first characteristic is a bond’s term—the length of time until the bond matures. Some bonds have short terms, such as a few months, while others have terms as long as 30 years. (The British government has even issued a bond that never matures, called a perpetuity. This bond pays interest forever, but the principal is never repaid.) The interest rate on a bond depends, in part, on its term. Long-term bonds are riskier than short-term bonds because holders of long-term bonds have to wait longer for repayment of principal. If a holder of a long-term bond needs his money earlier than the distant date of maturity, he has no choice but to sell the bond to someone else, perhaps at a reduced price. To compensate for this risk, long-term bonds usually pay higher interest rates than short-term bonds.

The second important characteristic of a bond is its credit risk—the probability that the borrower will fail to pay some of the interest or principal. Such a failure to pay is called a default. Borrowers can (and sometimes do) default on their loans by declaring bankruptcy. When bond buyers perceive that the probability of default is high, they demand a higher interest rate to compensate them for this risk. Because the U.S. government is considered a safe credit risk, government bonds tend to pay low interest rates. By contrast, financially shaky corporations raise money by issuing junk bonds, which pay very high interest rates. Buyers of bonds can judge credit risk by checking with various private agencies, such as Standard & Poor’s, which rate the credit risk of different bonds.

The third important characteristic of a bond is its tax treatment—the way the tax laws treat the interest earned on the bond. The interest on most bonds is taxable income so that the bond owner has to pay a portion of the interest in income taxes. By contrast, when state and local governments issue bonds, called municipal bonds, the bond owners are not required to pay federal income tax on the interest income. Because of this tax advantage, bonds issued by state and local governments pay a lower interest rate than bonds issued by corporations or the federal government.

The Stock Market Another way for Intel to raise funds to build a new semiconductor factory is to sell stock in the company. Stock represents ownership in a firm and, therefore, a claim to the profits that the firm makes. For example, if Intel sells a total of 1,000,000 shares of stock, then each share represents ownership of 1/1,000,000 of the business.

The sale of stock to raise money is called equity finance, whereas the sale of bonds is called debt finance. Although corporations use both equity and debt finance to raise money for new investments, stocks and bonds are very different.
The owner of shares of Intel stock is a part owner of Intel, while the owner of an Intel bond is a creditor of the corporation. If Intel is very profitable, the stockholders enjoy the benefits of these profits, whereas the bondholders get only the interest on their bonds. And if Intel runs into financial difficulty, the bondholders are paid what they are due before stockholders receive anything at all. Compared to bonds, stocks offer the holder both higher risk and potentially higher return.

After a corporation issues stock by selling shares to the public, these shares trade among stockholders on organized stock exchanges. In these transactions, the corporation itself receives no money when its stock changes hands. The most important stock exchanges in the U.S. economy are the New York Stock Exchange, the American Stock Exchange, and NASDAQ (National Association of Securities Dealers Automated Quotation system). Most of the world’s countries have their own stock exchanges on which the shares of local companies trade.

The prices at which shares trade on stock exchanges are determined by the supply of and demand for the stock in these companies. Because stock represents ownership in a corporation, the demand for a stock (and thus its price) reflects people’s perception of the corporation’s future profitability. When people become optimistic about a company’s future, they raise their demand for its stock and thereby bid up the price of a share of stock. Conversely, when people come to expect a company to have little profit or even losses, the price of a share falls.

Various stock indexes are available to monitor the overall level of stock prices. A *stock index* is computed as an average of a group of stock prices. The most famous stock index is the Dow Jones Industrial Average, which has been computed regularly since 1896. It is now based on the prices of the stocks of thirty major U.S. companies, such as General Motors, General Electric, Microsoft, Coca-Cola, AT&T, and IBM. Another well-known stock index is the Standard & Poor’s 500 Index, which is based on the prices of the stocks of 500 major companies. Because stock prices reflect expected profitability, these stock indexes are watched closely as possible indicators of future economic conditions.

**FINANCIAL INTERMEDIARIES**

Financial intermediaries are financial institutions through which savers can indirectly provide funds to borrowers. The term *intermediary* reflects the role of these institutions in standing between savers and borrowers. Here we consider two of the most important financial intermediaries: banks and mutual funds.

**Banks**  If the owner of a small grocery store wants to finance an expansion of his business, he probably takes a strategy quite different from that of Intel. Unlike Intel, a small grocer would find it difficult to raise funds in the bond and stock markets. Most buyers of stocks and bonds prefer to buy those issued by larger, more familiar companies. The small grocer, therefore, most likely finances his business expansion with a loan from a local bank.

Banks are the financial intermediaries with which people are most familiar. A primary job of banks is to take in deposits from people who want to save and use these deposits to make loans to people who want to borrow. Banks pay depositors interest on their deposits and charge borrowers slightly higher interest on their loans. The difference between these rates of interest covers the banks’ costs and returns some profit to the owners of the banks.

Besides being financial intermediaries, banks play a second important role in the economy: They facilitate purchases of goods and services by allowing people
to write checks against their deposits. In other words, banks help create a special asset that people can use as a medium of exchange. A medium of exchange is an item that people can easily use to engage in transactions. A bank’s role in providing a medium of exchange distinguishes it from many other financial institutions. Stocks and bonds, like bank deposits, are a possible store of value for the wealth that people have accumulated in past saving, but access to this wealth is not as easy, cheap, and immediate as just writing a check. For now, we ignore this second role of banks, but we will return to it when we discuss the monetary system later in the book.

**Mutual Funds**  A financial intermediary of increasing importance in the U.S. economy is the mutual fund. A mutual fund is an institution that sells shares to the public and uses the proceeds to buy a selection, or portfolio, of various types of stocks, bonds, or both stocks and bonds. The shareholder of the mutual fund accepts all the risk and return associated with the portfolio. If the value of the portfolio rises, the shareholder benefits; if the value of the portfolio falls, the shareholder suffers the loss.
The primary advantage of mutual funds is that they allow people with small amounts of money to diversify. Buyers of stocks and bonds are well advised to heed the adage: Don’t put all your eggs in one basket. Because the value of any single stock or bond is tied to the fortunes of one company, holding a single kind of stock or bond is very risky. By contrast, people who hold a diverse portfolio of stocks and bonds face less risk because they have only a small stake in each company. Mutual funds make this diversification easy. With only a few hundred dollars, a person can buy shares in a mutual fund and, indirectly, become the part owner or creditor of hundreds of major companies. For this service, the company operating the mutual fund charges shareholders a fee, usually between 0.5 and 2.0 percent of assets each year.

A second advantage claimed by mutual fund companies is that mutual funds give ordinary people access to the skills of professional money managers. The managers of most mutual funds pay close attention to the developments and prospects of the companies in which they buy stock. These managers buy the stock of companies they view as having a profitable future and sell the stock of companies with less promising prospects. This professional management, it is argued, should increase the return that mutual fund depositors earn on their savings.

Financial economists, however, are often skeptical of this second argument. With thousands of money managers paying close attention to each company’s prospects, the price of a company’s stock is usually a good reflection of the company’s true value. As a result, it is hard to “beat the market” by buying good stocks and selling bad ones. In fact, mutual funds called index funds, which buy all the stocks in a given stock index, perform somewhat better on average than mutual funds that take advantage of active trading by professional money managers. The explanation for the superior performance of index funds is that they keep costs low by buying and selling very rarely and by not having to pay the salaries of the professional money managers.

**Summing Up**

The U.S. economy contains a large variety of financial institutions. In addition to the bond market, the stock market, banks, and mutual funds, there are also pension funds, credit unions, insurance companies, and even the local loan shark. These institutions differ in many ways. When analyzing the macroeconomic role of the financial system, however, it is more important to keep in mind the similarity of these institutions than the differences. These financial institutions all serve the same goal: directing the resources of savers into the hands of borrowers.
QUICK QUIZ  What is stock? What is a bond? How are they different? How are they similar?

SAVING AND INVESTMENT IN THE NATIONAL INCOME ACCOUNTS

Events that occur within the financial system are central to understanding developments in the overall economy. As we have just seen, the institutions that make up this system—the bond market, the stock market, banks, and mutual funds—have the role of coordinating the economy’s saving and investment. And as we saw in the previous chapter, saving and investment are important determinants of long-run growth in GDP and living standards. As a result, macroeconomists need to understand how financial markets work and how various events and policies affect them.

As a starting point for an analysis of financial markets, we discuss in this section the key macroeconomic variables that measure activity in these markets. Our emphasis here is not on behavior but on accounting. Accounting refers to how various numbers are defined and added up. A personal accountant might help an individual add up her income and expenses. A national income accountant does the same thing for the economy as a whole. The national income accounts include, in particular, GDP and the many related statistics.

The rules of national income accounting include several important identities. Recall that an identity is an equation that must be true because of the way the variables in the equation are defined. Identities are useful to keep in mind, for they clarify how different variables are related to one another. Here we consider some accounting identities that shed light on the macroeconomic role of financial markets.

SOME IMPORTANT IDENTITIES

Recall that gross domestic product (GDP) is both total income in an economy and the total expenditure on the economy’s output of goods and services. GDP (denoted as $Y$) is divided into four components of expenditure: consumption ($C$), investment ($I$), government purchases ($G$), and net exports ($NX$). We write

$$Y = C + I + G + NX.$$ 

This equation is an identity because every dollar of expenditure that shows up on the left side also shows up in one of the four components on the right side. Because of the way each of the variables is defined and measured, this equation must always hold.

In this chapter, we simplify our analysis by assuming that the economy we are examining is closed. A closed economy is one that does not interact with other economies. In particular, a closed economy does not engage in international trade in goods and services, nor does it engage in international borrowing and lending. Actual economies are open economies—that is, they interact with other economies around the world. Nonetheless, assuming a closed economy is a useful simplification with which we can learn some lessons that apply to all economies. Moreover,
this assumption applies perfectly to the world economy (for interplanetary trade
is not yet common).

Because a closed economy does not engage in international trade, imports and
exports are exactly zero. Therefore, net exports (NX) are also zero. In this case, we
can write

\[ Y = C + I + G. \]

This equation states that GDP is the sum of consumption, investment, and gov-
ernment purchases. Each unit of output sold in a closed economy is consumed,
invested, or bought by the government.

To see what this identity can tell us about financial markets, subtract C and G
from both sides of this equation. We obtain

\[ Y - C - G = I. \]

The left side of this equation \((Y - C - G)\) is the total income in the economy that
remains after paying for consumption and government purchases: This amount is
called \textbf{national saving}, or just \textbf{saving}, and is denoted \(S\). Substituting \(S\) for \(Y - C - G\), we can write the last equation as

\[ S = I. \]

This equation states that saving equals investment.

To understand the meaning of national saving, it is helpful to manipulate the
definition a bit more. Let \(T\) denote the amount that the government collects from
households in taxes minus the amount it pays back to households in the form
of transfer payments (such as Social Security and welfare). We can then write
national saving in either of two ways:

\[ S = Y - C - G \]

or

\[ S = (Y - T - C) + (T - G). \]

These equations are the same because the two \(T\)s in the second equation cancel
each other, but each reveals a different way of thinking about national saving. In
particular, the second equation separates national saving into two pieces: private
saving \((Y - T - C)\) and public saving \((T - G)\).

Consider each of these two pieces. \textbf{Private saving} is the amount of income that
households have left after paying their taxes and paying for their consumption.
In particular, because households receive income of \(Y\), pay taxes of \(T\), and spend
\(C\) on consumption, private saving is \(Y - T - C\). \textbf{Public saving} is the amount of tax
revenue that the government has left after paying for its spending. The govern-
ment receives \(T\) in tax revenue and spends \(G\) on goods and services. If \(T\) exceeds
\(G\), the government runs a \textbf{budget surplus} because it receives more money than it
spends. This surplus of \(T - G\) represents public saving. If the government spends
more than it receives in tax revenue, then \(G\) is larger than \(T\). In this case, the gov-
ernment runs a \textbf{budget deficit}, and public saving \(T - G\) is a negative number.
Now consider how these accounting identities are related to financial markets. The equation $S = I$ reveals an important fact: For the economy as a whole, saving must be equal to investment. Yet this fact raises some important questions: What mechanisms lie behind this identity? What coordinates those people who are deciding how much to save and those people who are deciding how much to invest? The answer is the financial system. The bond market, the stock market, banks, mutual funds, and other financial markets and intermediaries stand between the two sides of the $S = I$ equation. They take in the nation’s saving and direct it to the nation’s investment.

**The Meaning of Saving and Investment**

The terms *saving* and *investment* can sometimes be confusing. Most people use these terms casually and sometimes interchangeably. By contrast, the macroeconomists who put together the national income accounts use these terms carefully and distinctly.

Consider an example. Suppose that Larry earns more than he spends and deposits his unspent income in a bank or uses it to buy some stock or a bond from a corporation. Because Larry’s income exceeds his consumption, he adds to the nation’s saving. Larry might think of himself as “investing” his money, but a macroeconomist would call Larry’s act saving rather than investment.

In the language of macroeconomics, investment refers to the purchase of new capital, such as equipment or buildings. When Moe borrows from the bank to build himself a new house, he adds to the nation’s investment. (Remember, the purchase of a new house is the one form of household spending that is investment rather than consumption.) Similarly, when the Curly Corporation sells some stock and uses the proceeds to build a new factory, it also adds to the nation’s investment.

Although the accounting identity $S = I$ shows that saving and investment are equal for the economy as a whole, this does not have to be true for every individual household or firm. Larry’s saving can be greater than his investment, and he can deposit the excess in a bank. Moe’s saving can be less than his investment, and he can borrow the shortfall from a bank. Banks and other financial institutions make these individual differences between saving and investment possible by allowing one person’s saving to finance another person’s investment.

**Quick Quiz** Define *private saving*, *public saving*, *national saving*, and *investment*. How are they related?

**The Market for Loanable Funds**

Having discussed some of the important financial institutions in our economy and the macroeconomic role of these institutions, we are ready to build a model of financial markets. Our purpose in building this model is to explain how financial markets coordinate the economy’s saving and investment. The model also gives us a tool with which we can analyze various government policies that influence saving and investment.

To keep things simple, we assume that the economy has only one financial market, called the **market for loanable funds**. All savers go to this market to deposit
their saving, and all borrowers go to this market to take out their loans. Thus, the term loanable funds refers to all income that people have chosen to save and lend out, rather than use for their own consumption, and to the amount that investors have chosen to borrow to fund new investment projects. In the market for loanable funds, there is one interest rate, which is both the return to saving and the cost of borrowing.

The assumption of a single financial market, of course, is not literally true. As we have seen, the economy has many types of financial institutions. But as we discussed in Chapter 2, the art in building an economic model is simplifying the world in order to explain it. For our purposes here, we can ignore the diversity of financial institutions and assume that the economy has a single financial market.

**Supply and Demand for Loanable Funds**

The economy’s market for loanable funds, like other markets in the economy, is governed by supply and demand. To understand how the market for loanable funds operates, therefore, we first look at the sources of supply and demand in that market.

The supply of loanable funds comes from people who have some extra income they want to save and lend out. This lending can occur directly, such as when a household buys a bond from a firm, or it can occur indirectly, such as when a household makes a deposit in a bank, which in turn uses the funds to make loans. In both cases, saving is the source of the supply of loanable funds.

The demand for loanable funds comes from households and firms who wish to borrow to make investments. This demand includes families taking out mortgages to buy new homes. It also includes firms borrowing to buy new equipment or build factories. In both cases, investment is the source of the demand for loanable funds.

The interest rate is the price of a loan. It represents the amount that borrowers pay for loans and the amount that lenders receive on their saving. Because a high interest rate makes borrowing more expensive, the quantity of loanable funds demanded falls as the interest rate rises. Similarly, because a high interest rate makes saving more attractive, the quantity of loanable funds supplied rises as the interest rate rises. In other words, the demand curve for loanable funds slopes downward, and the supply curve for loanable funds slopes upward.

Figure 1 shows the interest rate that balances the supply and demand for loanable funds. In the equilibrium shown, the interest rate is 5 percent, and the quantity of loanable funds demanded and the quantity of loanable funds supplied both equal $1,200 billion.

The adjustment of the interest rate to the equilibrium level occurs for the usual reasons. If the interest rate were lower than the equilibrium level, the quantity of loanable funds supplied would be less than the quantity of loanable funds demanded. The resulting shortage of loanable funds would encourage lenders to raise the interest rate they charge. A higher interest rate would encourage saving (thereby increasing the quantity of loanable funds supplied) and discourage borrowing for investment (thereby decreasing the quantity of loanable funds demanded). Conversely, if the interest rate were higher than the equilibrium level, the quantity of loanable funds supplied would exceed the quantity of loanable funds demanded. As lenders competed for the scarce borrowers, interest rates would be driven down. In this way, the interest rate approaches the equilibrium level at which the supply and demand for loanable funds exactly balance.
Recall that economists distinguish between the real interest rate and the nominal interest rate. The nominal interest rate is the interest rate as usually reported—the monetary return to saving and the monetary cost of borrowing. The real interest rate is the nominal interest rate corrected for inflation; it equals the nominal interest rate minus the inflation rate. Because inflation erodes the value of money over time, the real interest rate more accurately reflects the real return to saving and the real cost of borrowing. Therefore, the supply and demand for loanable funds depend on the real (rather than nominal) interest rate, and the equilibrium in Figure 1 should be interpreted as determining the real interest rate in the economy. For the rest of this chapter, when you see the term interest rate, you should remember that we are talking about the real interest rate.

This model of the supply and demand for loanable funds shows that financial markets work much like other markets in the economy. In the market for milk, for instance, the price of milk adjusts so that the quantity of milk supplied balances the quantity of milk demanded. In this way, the invisible hand coordinates the behavior of dairy farmers and the behavior of milk drinkers. Once we realize that saving represents the supply of loanable funds and investment represents the demand, we can see how the invisible hand coordinates saving and investment. When the interest rate adjusts to balance supply and demand in the market for loanable funds, it coordinates the behavior of people who want to save (the suppliers of loanable funds) and the behavior of people who want to invest (the demanders of loanable funds).

We can now use this analysis of the market for loanable funds to examine various government policies that affect the economy’s saving and investment. Because this model is just supply and demand in a particular market, we analyze any policy using the three steps discussed in Chapter 4. First, we decide whether the policy shifts the supply curve or the demand curve. Second, we determine the direction of the shift. Third, we use the supply-and-demand diagram to see how the equilibrium changes.
POLICY 1: SAVING INCENTIVES

American families save a smaller fraction of their incomes than their counterparts in many other countries, such as Japan and Germany. Although the reasons for these international differences are unclear, many U.S. policymakers view the low level of U.S. saving as a major problem. One of the Ten Principles of Economics in Chapter 1 is that a country’s standard of living depends on its ability to produce goods and services. And as we discussed in the preceding chapter, saving is an important long-run determinant of a nation’s productivity. If the United States could somehow raise its saving rate to the level that prevails in other countries, the growth rate of GDP would increase, and over time, U.S. citizens would enjoy a higher standard of living.

Another of the Ten Principles of Economics is that people respond to incentives. Many economists have used this principle to suggest that the low saving rate in the United States is at least partly attributable to tax laws that discourage saving. The U.S. federal government, as well as many state governments, collects revenue by taxing income, including interest and dividend income. To see the effects of this policy, consider a 25-year-old who saves $1,000 and buys a 30-year bond that pays an interest rate of 9 percent. In the absence of taxes, the $1,000 grows to $13,268 when the individual reaches age 55. Yet if that interest is taxed at a rate of, say, 33 percent, then the after-tax interest rate is only 6 percent. In this case, the $1,000 grows to only $5,743 after 30 years. The tax on interest income substantially reduces the future payoff from current saving and, as a result, reduces the incentive for people to save.

In response to this problem, many economists and lawmakers have proposed reforming the tax code to encourage greater saving. For example, one proposal is to expand eligibility for special accounts, such as Individual Retirement Accounts, that allow people to shelter some of their saving from taxation. Let’s consider the effect of such a saving incentive on the market for loanable funds, as illustrated in Figure 2. We analyze this policy following our three steps.

First, which curve would this policy affect? Because the tax change would alter the incentive for households to save at any given interest rate, it would affect the quantity of loanable funds supplied at each interest rate. Thus, the supply of loanable funds would shift. The demand for loanable funds would remain the same because the tax change would not directly affect the amount that borrowers want to borrow at any given interest rate.

Second, which way would the supply curve shift? Because saving would be taxed less heavily than under current law, households would increase their saving by consuming a smaller fraction of their income. Households would use this additional saving to increase their deposits in banks or to buy more bonds. The supply of loanable funds would increase, and the supply curve would shift to the right from $S_1$ to $S_2$, as shown in Figure 2.

Finally, we can compare the old and new equilibria. In the figure, the increased supply of loanable funds reduces the interest rate from 5 percent to 4 percent. The lower interest rate raises the quantity of loanable funds demanded from $1,200 billion to $1,600 billion. That is, the shift in the supply curve moves the market equilibrium along the demand curve. With a lower cost of borrowing, households and firms are motivated to borrow more to finance greater investment. Thus, if a reform of the tax laws encouraged greater saving, the result would be lower interest rates and greater investment.
Although this analysis of the effects of increased saving is widely accepted among economists, there is less consensus about what kinds of tax changes should be enacted. Many economists endorse tax reform aimed at increasing saving to stimulate investment and growth. Yet others are skeptical that these tax changes would have much effect on national saving. These skeptics also doubt the equity of the proposed reforms. They argue that, in many cases, the benefits of the tax changes would accrue primarily to the wealthy, who are least in need of tax relief.

**POLICY 2: INVESTMENT INCENTIVES**

Suppose that Congress passed a tax reform aimed at making investment more attractive. In essence, this is what Congress does when it institutes an investment tax credit, which it does from time to time. An investment tax credit gives a tax advantage to any firm building a new factory or buying a new piece of equipment. Let’s consider the effect of such a tax reform on the market for loanable funds, as illustrated in Figure 3.

First, would the law affect supply or demand? Because the tax credit would reward firms that borrow and invest in new capital, it would alter investment at any given interest rate and, thereby, change the demand for loanable funds. By contrast, because the tax credit would not affect the amount that households save at any given interest rate, it would not affect the supply of loanable funds.

Second, which way would the demand curve shift? Because firms would have an incentive to increase investment at any interest rate, the quantity of loanable funds demanded would be higher at any given interest rate. Thus, the demand curve for loanable funds would move to the right, as shown by the shift from $D_1$ to $D_2$ in the figure.
In Praise of Misers

In this opinion piece, economist Steven Landsburg defends Ebenezer Scrooge.

What I Like About Scrooge
By Steven E. Landsburg

Here's what I like about Ebenezer Scrooge: His meager lodgings were dark because darkness is cheap, and barely heated because coal is not free. His dinner was gruel, which he prepared himself. Scrooge paid no man to wait on him.

Scrooge has been called ungenerous. I say that’s a bum rap. What could be more generous than keeping your lamps unlit and your plate unfilled, leaving more fuel for others to burn and more food for others to eat? Who is a more benevolent neighbor than the man who employs no servants, freeing them to wait on someone else?

Oh, it might be slightly more complicated than that. Maybe when Scrooge demands less coal for his fire, less coal ends up being mined. But that’s fine, too. Instead of digging coal for Scrooge, some would-be miner is now free to perform some other service for himself or someone else.

Dickens tells us that the Lord Mayor, in the stronghold of the mighty Mansion House, gave orders to his 50 cooks and butlers to keep Christmas as a Lord Mayor’s household should—presumably for a houseful of guests who lavishly praised his generosity. The bricks, mortar, and labor that built the Mansion House might otherwise have built housing for hundreds; Scrooge, by living in three sparse rooms, deprived no man of a home. By employing no cooks or butlers, he ensured that cooks and butlers were available to some other household where guests reveled in ignorance of their debt to Ebenezer Scrooge.

In this whole world, there is nobody more generous than the miser—the man who could deplete the world’s resources but chooses not to. The only difference between miserliness and philanthropy is that the philanthropist serves a favored few while the miser spreads his largess far and wide.

If you build a house and refuse to buy a house, the rest of the world is one house richer. If you earn a dollar and refuse to spend a dollar, the rest of the world is one dollar richer—because you produced a dollar’s worth of goods and didn’t consume them.

Who exactly gets those goods? That depends on how you save. Put a dollar in the bank and you’ll bid down the interest rate by just enough so someone somewhere can afford an extra dollar’s worth of vacation or home improvement. Put a dollar in your mattress and (by effectively reducing the money supply) you’ll drive down prices by just enough so someone somewhere can have an extra dollar’s worth of coffee with his dinner. Scrooge, no doubt a canny investor, lent his money at interest. His less conventional namesake Scrooge McDuck filled a vault with dollar bills to roll around in. No matter. Ebenezer Scrooge lowered interest rates. Scrooge McDuck lowered prices. Each Scrooge enriched his neighbors as much as any Lord Mayor who invited the town in for a Christmas meal.

Saving is philanthropy, and—because this is both the Christmas season and the season of tax reform—it’s worth mentioning that the tax system should recognize as much. If there’s a tax deduction for charitable giving, there should be a tax deduction for saving. What you earn and don’t spend is your contribution to the world, and it’s equally a contribution whether you give it away or squirrel it away.

Of course, there’s always the threat that some meddling ghosts will come along and convince you to deplete your savings, at which point it makes sense (insofar as the taxation of income ever makes sense) to start taxing you. Which is exactly what individual retirement accounts are all about: They shield your earnings from taxation for as long as you save (that is, for as long as you let others enjoy the fruits of your labor), but no longer.

Great artists are sometimes unaware of the deepest meanings in their own creations. Though Dickens might not have recognized it, the primary moral of A Christmas Carol is that there should be no limit on IRA contributions. This is quite independent of all the other reasons why the tax system should encourage saving (e.g., the salutary effects on economic growth).

If Christmas is the season of selflessness, then surely one of the great symbols of Christmas should be Ebenezer Scrooge—the old Scrooge, not the reformed one. It’s taxes, not misers, that need reforming.
Third, consider how the equilibrium would change. In Figure 3, the increased demand for loanable funds raises the interest rate from 5 percent to 6 percent, and the higher interest rate in turn increases the quantity of loanable funds supplied from $1,200 billion to $1,400 billion, as households respond by increasing the amount they save. This change in household behavior is represented here as a movement along the supply curve. Thus, if a reform of the tax laws encouraged greater investment, the result would be higher interest rates and greater saving.

**POLICY 3: GOVERNMENT BUDGET DEFICITS AND SURPLUSES**

A perpetual topic of political debate is the status of the government budget. Recall that a **budget deficit** is an excess of government spending over tax revenue. Governments finance budget deficits by borrowing in the bond market, and the accumulation of past government borrowing is called the **government debt**. A **budget surplus**, an excess of tax revenue over government spending, can be used to repay some of the government debt. If government spending exactly equals tax revenue, the government is said to have a **balanced budget**.

Imagine that the government starts with a balanced budget and then, because of a tax cut or a spending increase, starts running a budget deficit. We can analyze the effects of the budget deficit by following our three steps in the market for loanable funds, as illustrated in Figure 4.

First, which curve shifts when the government starts running a budget deficit? Recall that national saving—the source of the supply of loanable funds—is composed of private saving and public saving. A change in the government budget
balance represents a change in public saving and, thereby, in the supply of loanable funds. Because the budget deficit does not influence the amount that households and firms want to borrow to finance investment at any given interest rate, it does not alter the demand for loanable funds.

Second, which way does the supply curve shift? When the government runs a budget deficit, public saving is negative, and this reduces national saving. In other words, when the government borrows to finance its budget deficit, it reduces the supply of loanable funds available to finance investment by households and firms. Thus, a budget deficit shifts the supply curve for loanable funds to the left from \( S_1 \) to \( S_2 \), as shown in Figure 4.

Third, we can compare the old and new equilibria. In the figure, when the budget deficit reduces the supply of loanable funds, the interest rate rises from 5 percent to 6 percent. This higher interest rate then alters the behavior of the households and firms that participate in the loan market. In particular, many demanders of loanable funds are discouraged by the higher interest rate. Fewer families buy new homes, and fewer firms choose to build new factories. The fall in investment because of government borrowing is called crowding out and is represented in the figure by the movement along the demand curve from a quantity of $1,200 billion in loanable funds to a quantity of $800 billion. That is, when the government borrows to finance its budget deficit, it crowds out private borrowers who are trying to finance investment.
Thus, the most basic lesson about budget deficits follows directly from their effects on the supply and demand for loanable funds: When the government reduces national saving by running a budget deficit, the interest rate rises, and investment falls. Because investment is important for long-run economic growth, government budget deficits reduce the economy’s growth rate.

Why, you might ask, does a budget deficit affect the supply of loanable funds, rather than the demand for them? After all, the government finances a budget deficit by selling bonds, thereby borrowing from the private sector. Why does increased borrowing from the government shift the supply curve, while increased borrowing by private investors shifts the demand curve? To answer this question, we need to examine more precisely the meaning of “loanable funds.” The model as presented here takes this term to mean the flow of resources available to fund private investment; thus, a government budget deficit reduces the supply of loanable funds. If, instead, we had defined the term “loanable funds” to mean the flow of resources available from private saving, then the government budget deficit would increase demand rather than reduce supply. Changing the interpretation of the term would cause a semantic change in how we described the model, but the bottom line from the analysis would be the same: In either case, a budget deficit increases the interest rate, thereby crowding out private borrowers who are relying on financial markets to fund private investment projects.

Now that we understand the impact of budget deficits, we can turn the analysis around and see that government budget surpluses have the opposite effects. When government collects more in tax revenue than it spends, it saves the difference by retiring some of the outstanding government debt. This budget surplus, or public saving, contributes to national saving. Thus, a budget surplus increases the supply of loanable funds, reduces the interest rate, and stimulates investment. Higher investment, in turn, means greater capital accumulation and more rapid economic growth.

THE HISTORY OF U.S. GOVERNMENT DEBT

How indebted is the U.S. government? The answer to this question varies substantially over time. Figure 5 shows the debt of the U.S. federal government expressed as a percentage of U.S. GDP. It shows that the government debt has fluctuated from zero in 1836 to 107 percent of GDP in 1945. In recent years, government debt has been between 30 and 40 percent of GDP.

The behavior of the debt-GDP ratio is one gauge of what’s happening with the government’s finances. Because GDP is a rough measure of the government’s tax base, a declining debt-GDP ratio indicates that the government indebtedness is shrinking relative to its ability to raise tax revenue. This suggests that the government is, in some sense, living within its means. By contrast, a rising debt-GDP ratio means that the government indebtedness is increasing relative to its ability to raise tax revenue. It is often interpreted as meaning that fiscal policy—government spending and taxes—cannot be sustained forever at current levels.

Throughout history, the primary cause of fluctuations in government debt is war. When wars occur, government spending on national defense rises substantially to pay for soldiers and military equipment. Taxes sometimes rise as well but typically by much less than the increase in spending. The result is a budget deficit and increasing government debt. When the war is over, government spending declines, and the debt-GDP ratio starts declining as well.
There are two reasons to believe that debt financing of war is an appropriate policy. First, it allows the government to keep tax rates smooth over time. Without debt financing, tax rates would have to rise sharply during wars, and this would cause a substantial decline in economic efficiency. Second, debt financing of wars shifts part of the cost of wars to future generations, who will have to pay off the government debt. This is arguably a fair distribution of the burden, for future generations get some of the benefit when one generation fights a war to defend the nation against foreign aggressors.

One large increase in government debt that cannot be explained by war is the increase that occurred beginning around 1980. When President Ronald Reagan took office in 1981, he was committed to smaller government and lower taxes. Yet he found cutting government spending to be more difficult politically than cutting taxes. The result was the beginning of a period of large budget deficits that continued not only through Reagan’s time in office but also for many years thereafter. As a result, government debt rose from 26 percent of GDP in 1980 to 50 percent of GDP in 1993.

As we discussed earlier, government budget deficits reduce national saving, investment, and long-run economic growth, and this is precisely why the rise in
government debt during the 1980s troubled many economists and policymakers. When Bill Clinton moved into the Oval Office in 1993, deficit reduction was his first major goal. Similarly, when the Republicans took control of Congress in 1995, deficit reduction was high on their legislative agenda. Both of these efforts substantially reduced the size of the government budget deficit, and it eventually turned into a surplus. As a result, by the late 1990s, the debt-GDP ratio was declining.

The debt-GDP ratio started rising again during the first few years of the George W. Bush presidency, as the budget surplus turned into a budget deficit. There were three reasons for this change. First, President Bush signed into law several major tax cuts, which he had promised during the 2000 presidential campaign. Second, in 2001, the economy experienced a recession (a reduction in economic activity), which automatically decreased tax revenue and increased government spending. Third, the war on terrorism following the September 11 attacks and then the war in Iraq led to increases in government spending.

**QUICK QUIZ** If more Americans adopted a “live for today” approach to life, how would this affect saving, investment, and the interest rate?

**CONCLUSION**

“Neither a borrower nor a lender be,” Polonius advises his son in Shakespeare’s Hamlet. If everyone followed this advice, this chapter would have been unnecessary.

Few economists would agree with Polonius. In our economy, people borrow and lend often, and usually for good reason. You may borrow one day to start your own business or to buy a home. And people may lend to you in the hope that the interest you pay will allow them to enjoy a more prosperous retirement. The financial system has the job of coordinating all this borrowing and lending activity.

In many ways, financial markets are like other markets in the economy. The price of loanable funds—the interest rate—is governed by the forces of supply and demand, just as other prices in the economy are. And we can analyze shifts in supply or demand in financial markets as we do in other markets. One of the Ten Principles of Economics introduced in Chapter 1 is that markets are usually a good way to organize economic activity. This principle applies to financial markets as well. When financial markets bring the supply and demand for loanable funds into balance, they help allocate the economy’s scarce resources to their most efficient use.

In one way, however, financial markets are special. Financial markets, unlike most other markets, serve the important role of linking the present and the future. Those who supply loanable funds—savers—do so because they want to convert some of their current income into future purchasing power. Those who demand loanable funds—borrowers—do so because they want to invest today in order to have additional capital in the future to produce goods and services. Thus, well-functioning financial markets are important not only for current generations but also for future generations who will inherit many of the resulting benefits.
SUMMARY

- The U.S. financial system is made up of many types of financial institutions, such as the bond market, the stock market, banks, and mutual funds. All these institutions act to direct the resources of households that want to save some of their income into the hands of households and firms that want to borrow.

- National income accounting identities reveal some important relationships among macroeconomic variables. In particular, for a closed economy, national saving must equal investment. Financial institutions are the mechanism through which the economy matches one person’s saving with another person’s investment.

- The interest rate is determined by the supply and demand for loanable funds. The supply of loanable funds comes from households that want to save some of their income and lend it out. The demand for loanable funds comes from households and firms that want to borrow for investment. To analyze how any policy or event affects the interest rate, one must consider how it affects the supply and demand for loanable funds.

- National saving equals private saving plus public saving. A government budget deficit represents negative public saving and, therefore, reduces national saving and the supply of loanable funds available to finance investment. When a government budget deficit crowds out investment, it reduces the growth of productivity and GDP.

KEY CONCEPTS

- financial system, p. 392
- financial markets, p. 392
- bond, p. 392
- stock, p. 393
- financial intermediaries, p. 394
- mutual fund, p. 395
- national saving (saving), p. 398
- private saving, p. 398
- public saving, p. 398
- budget surplus, p. 398
- budget deficit, p. 398
- market for loanable funds, p. 399
- crowding out, p. 406

QUESTIONS FOR REVIEW

1. What is the role of the financial system? Name and describe two markets that are part of the financial system in the U.S. economy. Name and describe two financial intermediaries.

2. Why is it important for people who own stocks and bonds to diversify their holdings? What type of financial institution makes diversification easier?

3. What is national saving? What is private saving? What is public saving? How are these three variables related?

4. What is investment? How is it related to national saving?

5. Describe a change in the tax code that might increase private saving. If this policy were implemented, how would it affect the market for loanable funds?

6. What is a government budget deficit? How does it affect interest rates, investment, and economic growth?
1. For each of the following pairs, which bond would you expect to pay a higher interest rate? Explain.
   a. a bond of the U.S. government or a bond of an East European government
   b. a bond that repays the principal in year 2013 or a bond that repays the principal in year 2030
   c. a bond from Coca-Cola or a bond from a software company you run in your garage
   d. a bond issued by the federal government or a bond issued by New York State
2. Theodore Roosevelt once said, “There is no moral difference between gambling at cards or in lotteries or on the race track and gambling in the stock market.” What social purpose do you think is served by the existence of the stock market?
3. When the Russian government defaulted on its debt to foreigners in 1998, interest rates rose on bonds issued by many other developing countries. Why do you suppose this happened?
4. Many workers hold large amounts of stock issued by the firms at which they work. Why do you suppose companies encourage this behavior? Why might a person not want to hold stock in the company where he works?
5. Explain the difference between saving and investment as defined by a macroeconomist. Which of the following situations represent investment? Saving? Explain.
   a. Your family takes out a mortgage and buys a new house.
   b. You use your $200 paycheck to buy stock in AT&T.
   c. Your roommate earns $100 and deposits it in her account at a bank.
   d. You borrow $1,000 from a bank to buy a car to use in your pizza delivery business.
6. Suppose GDP is $8 trillion, taxes are $1.5 trillion, private saving is $0.5 trillion, and public saving is $0.2 trillion. Assuming this economy is closed, calculate consumption, government purchases, national saving, and investment.
7. Economists in Funlandia, a closed economy, have collected the following information about the economy for a particular year:
   \[ Y = 10,000 \]
   \[ C = 6,000 \]
   \[ T = 1,500 \]
   \[ G = 1,700 \]
   The economists also estimate that the investment function is:
   \[ I = 3,300 - 100r \]
   where \( r \) is the country’s real interest rate, expressed as a percentage. Calculate private saving, public saving, national saving, investment, and the equilibrium real interest rate.
8. Suppose that Intel is considering building a new chip-making factory.
   a. Assuming that Intel needs to borrow money in the bond market, why would an increase in interest rates affect Intel’s decision about whether to build the factory?
   b. If Intel has enough of its own funds to finance the new factory without borrowing, would an increase in interest rates still affect Intel’s decision about whether to build the factory? Explain.
9. Suppose the government borrows $20 billion more next year than this year.
   a. Use a supply-and-demand diagram to analyze this policy. Does the interest rate rise or fall?
   b. What happens to investment? To private saving? To public saving? To national saving? Compare the size of the changes to the $20 billion of extra government borrowing.
   c. How does the elasticity of supply of loanable funds affect the size of these changes?
   d. How does the elasticity of demand for loanable funds affect the size of these changes?
   e. Suppose households believe that greater government borrowing today implies higher taxes to pay off the government debt in the future. What does this belief do to private
11. This chapter explains that investment can be increased both by reducing taxes on private saving and by reducing the government budget deficit.

a. Why is it difficult to implement both of these policies at the same time?

b. What would you need to know about private saving to judge which of these two policies would be a more effective way to raise investment?

10. “Some economists worry that the aging populations of industrial countries are going to start running down their savings just when the investment appetite of emerging economies is growing” (Economist, May 6, 1995). Illustrate the effect of these phenomena on the world market for loanable funds.
The Basic Tools of Finance

S

ometimes in your life, you will have to deal with the economy’s financial system. You will deposit your savings in a bank account, or you will take out a mortgage to buy a house. After you take a job, you will decide whether to invest your retirement account in stocks, bonds, or other financial instruments. You may try to put together your own stock portfolio, and then you will have to decide between betting on established companies such as General Electric or newer ones such as Google. And whenever you watch the evening news, you will hear reports about whether the stock market is up or down, together with the often feeble attempts to explain why the market behaves as it does.

If you reflect for a moment on the many financial decisions you will make during your life, you will see two related elements in almost all of them: time and risk. As we saw in the preceding two chapters, the financial system coordinates the economy’s saving and investment, which in turn are crucial determinants of economic growth. Most fundamentally, the financial system concerns decisions and actions we undertake today that will affect our lives in the future. But the future is unknown. When a person decides to allocate some saving, or a firm decides to undertake an investment, the decision is based on a guess about the likely result. The actual result, however, could end up being very different from what was expected.

This chapter introduces some tools that help us understand the decisions that people make as they participate in financial markets. The field of finance develops...
these tools in great detail, and you may choose to take courses that focus on this topic. But because the financial system is so important to the functioning of the economy, many of the basic insights of finance are central to understanding how the economy works. The tools of finance may also help you think through some of the decisions that you will make in your own life.

This chapter takes up three topics. First, we discuss how to compare sums of money at different points in time. Second, we discuss how to manage risk. Third, we build on our analysis of time and risk to examine what determines the value of an asset, such as a share of stock.

**PRESENT VALUE: MEASURING THE TIME VALUE OF MONEY**

Imagine that someone offers to give you $100 today or $100 in 10 years. Which would you choose? This is an easy question. Getting $100 today is better because you can always deposit the money in a bank, still have it in 10 years, and earn interest on the $100 along the way. The lesson: Money today is more valuable than the same amount of money in the future.

Now consider a harder question: Imagine that someone offers you $100 today or $200 in 10 years. Which would you choose? To answer this question, you need some way to compare sums of money from different points in time. Economists do this with a concept called present value. The present value of any future sum of money is the amount today that would be needed, at current interest rates, to produce that future sum.

To learn how to use the concept of present value, let’s work through a couple of simple examples:

**Question:** If you put $100 in a bank account today, how much will it be worth in \( N \) years? That is, what will be the future value of this $100?

**Answer:** Let’s use \( r \) to denote the interest rate expressed in decimal form (so an interest rate of 5 percent means \( r = 0.05 \)). Suppose that interest is paid annually and that the interest paid remains in the bank account to earn more interest—a process called compounding. Then the $100 will become

\[
(1 + r) \times 100 \quad \text{after 1 year,}
\]

\[
(1 + r) \times (1 + r) \times 100 = (1 + r)^2 \times 100 \quad \text{after 2 years,}
\]

\[
(1 + r) \times (1 + r) \times (1 + r) \times 100 = (1 + r)^3 \times 100 \quad \text{after 3 years, \ldots}
\]

\[
(1 + r)^N \times 100 \quad \text{after \( N \) years.}
\]

For example, if we are investing at an interest rate of 5 percent for 10 years, then the future value of the $100 will be \((1.05)^{10} \times 100\), which is $163.

**Question:** Now suppose you are going to be paid $200 in \( N \) years. What is the present value of this future payment? That is, how much would you have to deposit in a bank right now to yield $200 in \( N \) years?

...
Answer: To answer this question, just turn the previous answer on its head. In the last question, we computed a future value from a present value by multiplying by the factor \((1 + r)^N\). To compute a present value from a future value, we divide by the factor \((1 + r)^N\). Thus, the present value of $200 in \(N\) years is \(\frac{200}{(1 + r)^N}\). If that amount is deposited in a bank today, after \(N\) years it would become \((1 + r)^N \times \left(\frac{200}{(1 + r)^N}\right)\), which is $200. For instance, if the interest rate is 5 percent, the present value of $200 in 10 years is $200/(1.05)^{10}$, which is $123. This means that $123 deposited today in a bank account that earned 5 percent would produce $200 after 10 years.

This illustrates the general formula:

- If \(r\) is the interest rate, then an amount \(X\) to be received in \(N\) years has a present value of \(\frac{X}{(1 + r)^N}\).

Because the possibility of earning interest reduces the present value below the amount \(X\), the process of finding a present value of a future sum of money is called discounting. This formula shows precisely how much future sums should be discounted.

Let's now return to our earlier question: Should you choose $100 today or $200 in 10 years? We can infer from our calculation of present value that if the interest rate is 5 percent, you should prefer the $200 in 10 years. The future $200 has a present value of $123, which is greater than $100. You are better off waiting for the future sum.

Notice that the answer to our question depends on the interest rate. If the interest rate were 8 percent, then the $200 in 10 years would have a present value of $200/(1.08)^{10}$, which is only $93. In this case, you should take the $100 today. Why should the interest rate matter for your choice? The answer is that the higher the interest rate, the more you can earn by depositing your money in a bank, so the more attractive getting $100 today becomes.

The concept of present value is useful in many applications, including the decisions that companies face when evaluating investment projects. For instance, imagine that General Motors is thinking about building a new factory. Suppose that the factory will cost $100 million today and will yield the company $200 million in 10 years. Should General Motors undertake the project? You can see that this decision is exactly like the one we have been studying. To make its decision, the company will compare the present value of the $200 million return to the $100 million cost.

The company’s decision, therefore, will depend on the interest rate. If the interest rate is 5 percent, then the present value of the $200 million return from the factory is $123 million, and the company will choose to pay the $100 million cost. By contrast, if the interest rate is 8 percent, then the present value of the return is only $93 million, and the company will decide to forgo the project. Thus, the concept of present value helps explain why investment—and thus the quantity of loanable funds demanded—declines when the interest rate rises.

Here is another application of present value: Suppose you win a million-dollar lottery and are given a choice between $20,000 a year for 50 years (totaling $1,000,000) or an immediate payment of $400,000. Which would you choose? To make the right choice, you need to calculate the present value of the stream of
payments. Let’s suppose the interest rate is 7 percent. After performing 50 calculations similar to those above (one calculation for each payment) and adding up the results, you would learn that the present value of this million-dollar prize at a 7 percent interest rate is only $276,000. You are better off picking the immediate payment of $400,000. The million dollars may seem like more money, but the future cash flows, once discounted to the present, are worth far less.

**Quick Quiz** The interest rate is 7 percent. What is the present value of $150 to be received in 10 years?

**Managing Risk**

Life is full of gambles. When you go skiing, you risk breaking your leg in a fall. When you drive to work, you risk a car accident. When you put some of your savings in the stock market, you risk a fall in stock prices. The rational response to this risk is not necessarily to avoid it at any cost but to take it into account in your decision making. Let’s consider how a person might do that.
RISK AVERSION

Most people are risk averse. This means more than that people dislike bad things happening to them. It means that they dislike bad things more than they like comparable good things.

For example, suppose a friend offers you the following opportunity. He will toss a coin. If it comes up heads, he will pay you $1,000. But if it comes up tails, you will have to pay him $1,000. Would you accept the bargain? You wouldn’t if you were risk averse. For a risk-averse person, the pain of losing the $1,000 would exceed the pleasure from winning $1,000.

Economists have developed models of risk aversion using the concept of utility, which is a person’s subjective measure of well-being or satisfaction. Every level of wealth provides a certain amount of utility, as shown by the utility function in Figure 1. But the function exhibits the property of diminishing marginal utility: The more wealth a person has, the less utility he gets from an additional dollar. Thus, in the figure, the utility function gets flatter as wealth increases. Because of diminishing marginal utility, the utility lost from losing the $1,000 bet is more than the utility gained from winning it. As a result, people are risk averse.

Risk aversion provides the starting point for explaining various things we observe in the economy. Let’s consider three of them: insurance, diversification, and the risk-return trade-off.

THE MARKETS FOR INSURANCE

One way to deal with risk is to buy insurance. The general feature of insurance contracts is that a person facing a risk pays a fee to an insurance company, which in return agrees to accept all or part of the risk. There are many types of insurance. Car insurance covers the risk of your being in an auto accident, fire insurance covers the risk that your house will burn down, health insurance covers the risk of getting sick.

The Utility Function

This utility function shows how utility, a subjective measure of satisfaction, depends on wealth. As wealth rises, the utility function becomes flatter, reflecting the property of diminishing marginal utility. Because of diminishing marginal utility, a $1,000 loss decreases utility by more than a $1,000 gain increases it.
that you might need expensive medical treatment, and life insurance covers the risk that you will die and leave your family without your income. There is also insurance against the risk of living too long: For a fee paid today, an insurance company will pay you an annuity—a regular income every year until you die.

In a sense, every insurance contract is a gamble. It is possible that you will not be in an auto accident, that your house will not burn down, and that you will not need expensive medical treatment. In most years, you will pay the insurance company the premium and get nothing in return except peace of mind. Indeed, the insurance company is counting on the fact that most people will not make claims on their policies; otherwise, it couldn’t pay out the large claims to the unlucky few and still stay in business.

From the standpoint of the economy as a whole, the role of insurance is not to eliminate the risks inherent in life but to spread them around more efficiently.
Consider fire insurance, for instance. Owning fire insurance does not reduce the risk of losing your home in a fire. But if that unlucky event occurs, the insurance company compensates you. The risk, rather than being borne by you alone, is shared among the thousands of insurance-company shareholders. Because people are risk averse, it is easier for 10,000 people to bear 1/10,000 of the risk than for one person to bear the entire risk himself.

The markets for insurance suffer from two types of problems that impede their ability to spread risk. One problem is adverse selection: A high-risk person is more likely to apply for insurance than a low-risk person because a high-risk person would benefit more from insurance protection. A second problem is moral hazard: After people buy insurance, they have less incentive to be careful about their risky behavior because the insurance company will cover much of the resulting losses. Insurance companies are aware of these problems, but they cannot fully guard against them. An insurance company cannot perfectly distinguish between high-risk and low-risk customers, and it cannot monitor all of its customers’ risky behavior. The price of insurance reflects the actual risks that the insurance company will face after the insurance is bought. The high price of insurance is why some people, especially those who know themselves to be low-risk, decide against buying it and, instead, endure some of life’s uncertainty on their own.

**Diversification of Firm-Specific Risk**

In 2002, Enron, a large and once widely respected company, went bankrupt amid accusations of fraud and accounting irregularities. Several of the company’s top executives were prosecuted and ended up going to prison. The saddest part of the story, however, involved thousands of lower-level employees. Not only did they lose their jobs, but many lost their life savings as well. The employees had about two-thirds of their retirement funds in Enron stock, which became worthless.

If there is one piece of practical advice that finance offers to risk-averse people, it is this: “Don’t put all your eggs in one basket.” You may have heard this before, but finance has turned this folk wisdom into a science. It goes by the name diversification.

The market for insurance is one example of diversification. Imagine a town with 10,000 homeowners, each facing the risk of a house fire. If someone starts an insurance company and each person in town becomes both a shareholder and a policyholder of the company, they all reduce their risk through diversification. Each person now faces 1/10,000 of the risk of 10,000 possible fires, rather than the entire risk of a single fire in his own home. Unless the entire town catches fire at the same time, the downside that each person faces is much smaller.

When people use their savings to buy financial assets, they can also reduce risk through diversification. A person who buys stock in a company is placing a bet on the future profitability of that company. That bet is often quite risky because companies’ fortunes are hard to predict. Microsoft evolved from a start-up by some geeky teenagers to one of the world’s most valuable companies in only a few years; Enron went from one of the world’s most respected companies to an almost worthless one in only a few months. Fortunately, a shareholder need not tie his own fortune to that of any single company. Risk can be reduced by placing a large number of small bets, rather than a small number of large ones.

Figure 2 shows how the risk of a portfolio of stocks depends on the number of stocks in the portfolio. Risk is measured here with a statistic called the standard
Diversification Reduces Risk
This figure shows how the risk of a portfolio, measured here with a statistic called the standard deviation, depends on the number of stocks in the portfolio. The investor is assumed to put an equal percentage of his portfolio in each of the stocks. Increasing the number of stocks reduces, but does not eliminate, the amount of risk in a stock portfolio.


The trade-off between risk and return
One of the Ten Principles of Economics in Chapter 1 is that people face trade-offs. The trade-off that is most relevant for understanding financial decisions is the trade-off between risk and return.

As we have seen, there are risks inherent in holding stocks, even in a diversified portfolio. But risk-averse people are willing to accept this uncertainty because they are compensated for doing so. Historically, stocks have offered much higher rates of return than alternative financial assets, such as bonds and bank savings...
accounts. Over the past two centuries, stocks offered an average real return of about 8 percent per year, while short-term government bonds paid a real return of only 3 percent per year.

When deciding how to allocate their savings, people have to decide how much risk they are willing to undertake to earn a higher return. Figure 3 illustrates the risk-return trade-off for a person choosing how to allocate his portfolio between two asset classes:

- The first asset class is a diversified group of risky stocks, with an average return of 8 percent and a standard deviation of 20 percent. (You may recall from a math or statistics class that a normal random variable stays within two standard deviations of its average about 95 percent of the time. Thus, while actual returns are centered around 8 percent, they typically vary from a gain of 48 percent to a loss of 32 percent.)
- The second asset class is a safe alternative, with a return of 3 percent and a standard deviation of zero. The safe alternative can be either a bank savings account or a government bond.

Each point in this figure represents a particular allocation of a portfolio between risky stocks and the safe asset. The figure shows that the more a person puts into stocks, the greater is both the risk and the return.

Acknowledging the risk-return trade-off does not, by itself, tell us what a person should do. The choice of a particular combination of risk and return depends on a person’s risk aversion, which reflects a person’s own preferences. But it is important for stockholders to realize that the higher average return that they enjoy comes at the price of higher risk.

**QUICK QUIZ** Describe three ways that a risk-averse person might reduce the risk she faces.
Now that we have developed a basic understanding of the two building blocks of finance—time and risk—let’s apply this knowledge. This section considers a simple question: What determines the price of a share of stock? Like most prices, the answer is supply and demand. But that is not the end of the story. To understand stock prices, we need to think more deeply about what determines a person’s willingness to pay for a share of stock.

**Fundamental Analysis**

Let’s imagine that you have decided to put 60 percent of your savings into stock, and to achieve diversification, you have decided to buy 20 different stocks. If you open up the newspaper, you will find thousands of stocks listed. How should you pick the 20 for your portfolio?

When you buy stock, you are buying shares in a business. When deciding which businesses you want to own, it is natural to consider two things: the value of that share of the business and the price at which the shares are being sold. If the price is less than the value, the stock is said to be *undervalued*. If the price is more than the value, the stock is said to be *overvalued*. If the price and the value are equal, the stock is said to be *fairly valued*. When choosing 20 stocks for your portfolio, you should prefer undervalued stocks. In these cases, you are getting a bargain by paying less than the business is worth.

This is easier said than done. Learning the price is easy: You can just look it up in the newspaper. Determining the value of the business is the hard part. The term *fundamental analysis* refers to the detailed analysis of a company to determine its value. Many Wall Street firms hire stock analysts to conduct such fundamental analysis and offer advice about which stocks to buy.

The value of a stock to a stockholder is what he gets out of owning it, which includes the present value of the stream of dividend payments and the final sale price. Recall that *dividends* are the cash payments that a company makes to its shareholders. A company’s ability to pay dividends, as well as the value of the stock when the stockholder sells his shares, depends on the company’s ability to earn profits. Its profitability, in turn, depends on a large number of factors: the demand for its product, how much competition it faces, how much capital it has in place, whether its workers are unionized, how loyal its customers are, what kinds of government regulations and taxes it faces, and so on. The job of fundamental analysts is to take all these factors into account to determine how much a share of stock in the company is worth.

If you want to rely on fundamental analysis to pick a stock portfolio, there are three ways to do it. One way is to do all the necessary research yourself, such as by reading through companies’ annual reports. A second way is to rely on the advice of Wall Street analysts. A third way is to buy a mutual fund, which has a manager who conducts fundamental analysis and makes the decision for you.

**The Efficient Markets Hypothesis**

There is another way to choose 20 stocks for your portfolio: Pick them randomly by, for instance, putting the stock pages on your bulletin board and throwing...
darts at the page. This may sound crazy, but there is reason to believe that it won’t lead you too far astray. That reason is called the **efficient markets hypothesis**.

To understand this theory, the starting point is to acknowledge that each company listed on a major stock exchange is followed closely by many money managers, such as the individuals who run mutual funds. Every day, these managers monitor news stories and conduct fundamental analysis to try to determine the stock’s value. Their job is to buy a stock when its price falls below its value and to sell it when its price rises above its value.

The second piece to the efficient markets hypothesis is that the equilibrium of supply and demand sets the market price. This means that, at the market price, the number of shares being offered for sale exactly equals the number of shares that people want to buy. In other words, at the market price, the number of people who think the stock is overvalued exactly balances the number of people who think it’s undervalued. As judged by the typical person in the market, all stocks are fairly valued all the time.

According to this theory, the stock market exhibits **informational efficiency**: It reflects all available information about the value of the asset. Stock prices change when information changes. When good news about the company’s prospects becomes public, the value and the stock price both rise. When the company’s prospects deteriorate, the value and price both fall. But at any moment in time, the market price is the best guess of the company’s value based on available information.

One implication of the efficient markets hypothesis is that stock prices should follow a **random walk**. This means that the changes in stock prices are impossible to predict from available information. If, based on publicly available information, a person could predict that a stock price would rise by 10 percent tomorrow, then the stock market must be failing to incorporate that information today. According to this theory, the only thing that can move stock prices is news that changes the market’s perception of the company’s value. But news must be unpredictable—otherwise, it wouldn’t really be news. For the same reason, changes in stock prices should be unpredictable.

If the efficient markets hypothesis is correct, then there is little point in spending many hours studying the business page to decide which 20 stocks to add to your portfolio. If prices reflect all available information, no stock is a better buy than any other. The best you can do is buy a diversified portfolio.

### Case Study: Random Walks and Index Funds

The efficient markets hypothesis is a theory about how financial markets work. The theory is probably not completely true: As we discuss in the next section, there is reason to doubt that stockholders are always rational and that stock prices are informationally efficient at every moment. Nonetheless, the efficient markets hypothesis does much better as a description of the world than you might think.

There is much evidence that stock prices, even if not exactly a random walk, are very close to it. For example, you might be tempted to buy stocks that have recently risen and avoid stocks that have recently fallen (or perhaps just the opposite). But statistical studies have shown that following such trends (or bucking them) fails to outperform the market. The correlation between how well a stock does one year and how well it does the following year is almost exactly zero.
Some of the best evidence in favor of the efficient markets hypothesis comes from the performance of index funds. An index fund is a mutual fund that buys all the stocks in a given stock index. The performance of these funds can be compared with that of actively managed mutual funds, where a professional portfolio manager picks stocks based on extensive research and alleged expertise. In essence, an index fund buys all stocks, whereas active funds are supposed to buy only the best stocks.

In practice, active managers usually fail to beat index funds, and in fact, most of them do worse. For example, in the five years ending April 2008, 76 percent of stock mutual funds failed to beat a broadly based index fund holding all stocks traded on U.S. stock exchanges. Most active portfolio managers give a lower return than index funds because they trade more frequently, incurring more trading costs, and because they charge greater fees as compensation for their alleged expertise.

What about the 24 percent of managers who did beat the market? Perhaps they are smarter than average, or perhaps they were luckier. If you have 5,000 people flipping coins ten times, on average about five will flip ten heads; these five might claim an exceptional coin-flipping skill, but they would have trouble replicating the feat. Similarly, studies have shown that mutual fund managers with a history of superior performance usually fail to maintain it in subsequent periods.

Lessons from the Brain-Damaged Investor
By Jane Spencer

People with certain kinds of brain damage may make better investment decisions. That is the conclusion of a new study offering some compelling evidence that mixing emotion with investing can lead to bad outcomes.

By linking brain science to investment behavior, researchers concluded that people with an impaired ability to experience emotions could actually make better financial decisions than other people under certain circumstances. The research is part of a fast-growing interdisciplinary field called “neuroeconomics” that explores the role biology plays in economic decision making, by combining insights from cognitive neuroscience, psychology and economics. The study was published last month in the journal Psychological Science, and was conducted by a team of researchers from Carnegie Mellon University, the Stanford Graduate School of Business and the University of Iowa.

The 15 brain-damaged participants that were the focus of the study had normal IQs, and the areas of their brains responsible for logic and cognitive reasoning were intact. But they had lesions in the region of the brain that controls emotions, which inhibited their ability to experience basic feelings such as fear or anxiety. The lesions were due to a range of causes, including stroke and disease, but they impaired the participants’ emotional functioning in a similar manner.

The study suggests the participants’ lack of emotional responsiveness actually gave them an advantage when they played a simple investment game. The emotionally impaired players were more willing to take gambles that had high payoffs because they lacked fear. Players with undamaged brain wiring, however, were more cautious and

Neurofinance

New research is exploring the linkages among economics, psychology, and brain science.
The Wall Street Journal published an example of this phenomenon on January 3, 2008. The paper reported that of the many thousands of mutual funds sold to the public, only thirty-one beat the Standard & Poor’s 500 index in each of the 8 years from 1999 to 2006. A skeptic of the efficient markets hypothesis might think that, subsequently, these highly performing funds would offer a better-than-average place to invest. In 2007, however, only fourteen of these thirty-one outperformed the index—about what would be expected from sheer chance. Exceptional past performance appears to give little reason to expect future success.

The efficient markets hypothesis says that it is impossible to beat the market. The accumulation of many studies in financial markets confirms that beating the market is, at best, extremely difficult. Even if the efficient markets hypothesis is not an exact description of the world, it contains a large element of truth.

**Market Irrationality**

The efficient markets hypothesis assumes that people buying and selling stock rationally process the information they have about the stock’s underlying value. But is the stock market really that rational? Or do stock prices sometimes deviate from reasonable expectations of their true value?
There is a long tradition suggesting that fluctuations in stock prices are partly psychological. In the 1930s, economist John Maynard Keynes suggested that asset markets are driven by the “animal spirits” of investors—irrational waves of optimism and pessimism. In the 1990s, as the stock market soared to new heights, Fed Chairman Alan Greenspan questioned whether the boom reflected “irrational exuberance.” Stock prices did subsequently fall, but whether the exuberance of the 1990s was irrational given the information available at the time remains debatable. Whenever the price of an asset rises above what appears to be its fundamental value, the market is said to be experiencing a speculative bubble.

The possibility of speculative bubbles in the stock market arises in part because the value of the stock to a stockholder depends not only on the stream of dividend payments but also on the final sale price. Thus, a person might be willing to pay more than a stock is worth today if she expects another person to pay even more for it tomorrow. When you evaluate a stock, you have to estimate not only the value of the business but also what other people will think the business is worth in the future.

There is much debate among economists about the frequency and importance of departures from rational pricing. Believers in market irrationality point out (correctly) that the stock market often moves in ways that are hard to explain on the basis of news that might alter a rational valuation. Believers in the efficient markets hypothesis point out (correctly) that it is impossible to know the correct, rational valuation of a company, so one should not quickly jump to the conclusion that any particular valuation is irrational. Moreover, if the market were irrational, a rational person should be able to take advantage of this fact; yet as the previous case study discussed, beating the market is nearly impossible.

**QUICK QUIZ**  *Fortune* magazine regularly publishes a list of the “most respected” companies. According to the efficient markets hypothesis, if you restrict your stock portfolio to these companies, will you earn a better than average return? Explain.

**CONCLUSION**

This chapter has developed some of the basic tools that people should (and often do) use as they make financial decisions. The concept of present value reminds us that a dollar in the future is less valuable than a dollar today, and it gives us a way to compare sums of money at different points in time. The theory of risk management reminds us that the future is uncertain and that risk-averse people can take precautions to guard against this uncertainty. The study of asset valuation tells us that the stock price of any company should reflect its expected future profitability.

Although most of the tools of finance are well established, there is more controversy about the validity of the efficient markets hypothesis and whether stock prices are, in practice, rational estimates of a company’s true worth. Rational or not, the large movements in stock prices that we observe have important macroeconomic implications. Stock market fluctuations often go hand in hand with fluctuations in the economy more broadly. We will revisit the stock market when we study economic fluctuations later in the book.
**SUMMARY**

- Because savings can earn interest, a sum of money today is more valuable than the same sum of money in the future. A person can compare sums from different times using the concept of present value. The present value of any future sum is the amount that would be needed today, given prevailing interest rates, to produce that future sum.

- Because of diminishing marginal utility, most people are risk averse. Risk-averse people can reduce risk by buying insurance, diversifying their holdings, and choosing a portfolio with lower risk and lower return.

- The value of an asset equals the present value of the cash flows the owner will receive. For a share of stock, these cash flows include the stream of dividends and the final sale price. According to the efficient markets hypothesis, financial markets process available information rationally, so a stock price always equals the best estimate of the value of the underlying business. Some economists question the efficient markets hypothesis, however, and believe that irrational psychological factors also influence asset prices.

**KEY CONCEPTS**

- Finance, p. 413
- Present value, p. 414
- Future value, p. 414
- Compounding, p. 414
- Risk aversion, p. 417
- Diversification, p. 419
- Firm-specific risk, p. 420
- Market risk, p. 420
- Fundamental analysis, p. 422
- Efficient markets hypothesis, p. 423
- Informational efficiency, p. 423
- Random walk, p. 423

**QUESTIONS FOR REVIEW**

1. The interest rate is 7 percent. Use the concept of present value to compare $200 to be received in 10 years and $300 to be received in 20 years.

2. What benefit do people get from the market for insurance? What two problems impede the insurance company from working perfectly?

3. What is diversification? Does a stockholder get more diversification going from 1 to 10 stocks or going from 100 to 120 stocks?

4. Comparing stocks and government bonds, which has more risk? Which pays a higher average return?

5. What factors should a stock analyst think about in determining the value of a share of stock?

6. Describe the efficient markets hypothesis and give a piece of evidence consistent with this hypothesis.

7. Explain the view of those economists who are skeptical of the efficient markets hypothesis.

**PROBLEMS AND APPLICATIONS**

1. According to an old myth, Native Americans sold the island of Manhattan about 400 years ago for $24. If they had invested this amount at an interest rate of 7 percent per year, how much would they have today?

2. A company has an investment project that would cost $10 million today and yield a payoff of $15 million in 4 years.
   a. Should the firm undertake the project if the interest rate is 11 percent? 10 percent? 9 percent? 8 percent?
b. Can you figure out the exact cutoff for the interest rate between profitability and non-profitability?

3. Your bank account pays an interest rate of 8 percent. You are considering buying a share of stock in XYZ Corporation for $110. After 1, 2, and 3 years, it will pay a dividend of $5. You expect to sell the stock after 3 years for $120. Is XYZ a good investment? Support your answer with calculations.

4. For each of the following kinds of insurance, give an example of behavior that can be called moral hazard and another example of behavior that can be called adverse selection.
   a. health insurance
   b. car insurance

5. Imagine that the U.S. Congress, recognizing the importance of being well dressed, started giving preferential tax treatment to “clothing insurance.” Under this new type of insurance, you would pay the insurance company an annual premium, the insurance company would then pay for 80 percent of your clothing expenses (you pay the remaining 20 percent), and the tax laws would partly subsidize your insurance premiums.
   a. How would the existence of such insurance affect the amount of clothing that people buy? How would you evaluate this change in behavior from the standpoint of economic efficiency?
   b. Who would choose to buy clothing insurance?
   c. Suppose that the average person now spends $2,000 a year on clothes. Would clothing insurance cost more or less than $2,000? Explain.
   d. In your view, is this congressional action a good idea? How would you compare this idea with the current tax treatment of health insurance?

6. Imagine that you intend to buy a portfolio of ten stocks with some of your savings. Should the stocks be of companies in the same industry? Should the stocks be of companies located in the same country? Explain.

7. Which kind of stock would you expect to pay the higher average return: stock in an industry that is very sensitive to economic conditions (such as an automaker) or stock in an industry that is relatively insensitive to economic conditions (such as a water company)? Why?

8. A company faces two kinds of risk. A firm-specific risk is that a competitor might enter its market and take some of its customers. A market risk is that the economy might enter a recession, reducing sales. Which of these two risks would more likely cause the company’s shareholders to demand a higher return? Why?

9. You have two roommates who invest in the stock market.
   a. One roommate says that he buys stock only in companies that everyone believes will experience big increases in profits in the future. How do you suppose the price-earnings ratio of these companies compares to the price-earnings ratio of other companies? What might be the disadvantage of buying stock in these companies?
   b. Another roommate says he only buys stock in companies that are cheap, which he measures by a low price-earnings ratio. How do you suppose the earnings prospects of these companies compare to those of other companies? What might be the disadvantage of buying stock in these companies?

10. When company executives buy and sell stock based on private information they obtain as part of their jobs, they are engaged in insider trading.
   a. Give an example of inside information that might be useful for buying or selling stock.
   b. Those who trade stocks based on inside information usually earn very high rates of return. Does this fact violate the efficient markets hypothesis?
   c. Insider trading is illegal. Why do you suppose that is?

11. Find some information on an index fund (such as the Vanguard Total Stock Market Index, ticker symbol VTSMX). How has this fund performed compared with other stock mutual funds over the past 5 or 10 years? (Hint: One place to look for data on mutual funds is http://www.morningstar.com.) What do you learn from this comparison?
Unemployment

Losing a job can be the most distressing economic event in a person’s life. Most people rely on their labor earnings to maintain their standard of living, and many people also get a sense of personal accomplishment from working. A job loss means a lower living standard in the present, anxiety about the future, and reduced self-esteem. It is not surprising, therefore, that politicians campaigning for office often speak about how their proposed policies will help create jobs.

In previous chapters, we have seen some of the forces that determine the level and growth of a country’s standard of living. A country that saves and invests a high fraction of its income, for instance, enjoys more rapid growth in its capital stock and GDP than a similar country that saves and invests less. An even more obvious determinant of a country’s standard of living is the amount of unemployment it typically experiences. People who would like to work but cannot find a job are not contributing to the economy’s production of goods and services. Although some degree of unemployment is inevitable in a complex economy with thousands of firms and millions of workers, the amount of unemployment varies substantially over time and across countries. When a country keeps its workers as fully employed as possible, it achieves a higher level of GDP than it would if it left many of its workers standing idle.

This chapter begins our study of unemployment. The problem of unemployment is usefully divided into two categories: the long-run problem and the
short-run problem. The economy’s natural rate of unemployment refers to the amount of unemployment that the economy normally experiences. Cyclical unemployment refers to the year-to-year fluctuations in unemployment around its natural rate, and it is closely associated with the short-run ups and downs of economic activity. Cyclical unemployment has its own explanation, which we defer until we study short-run economic fluctuations later in this book. In this chapter, we discuss the determinants of an economy’s natural rate of unemployment. As we will see, the designation natural does not imply that this rate of unemployment is desirable. Nor does it imply that it is constant over time or impervious to economic policy. It merely means that this unemployment does not go away on its own even in the long run.

We begin the chapter by looking at some of the relevant facts that describe unemployment. In particular, we examine three questions: How does the government measure the economy’s rate of unemployment? What problems arise in interpreting the unemployment data? How long are the unemployed typically without work?

We then turn to the reasons economies always experience some unemployment and the ways in which policymakers can help the unemployed. We discuss four explanations for the economy’s natural rate of unemployment: job search, minimum-wage laws, unions, and efficiency wages. As we will see, long-run unemployment does not arise from a single problem that has a single solution. Instead, it reflects a variety of related problems. As a result, there is no easy way for policymakers to reduce the economy’s natural rate of unemployment and, at the same time, to alleviate the hardships experienced by the unemployed.

**IDENTIFYING UNEMPLOYMENT**

Let’s start by examining more precisely what the term unemployment means.

**HOW IS UNEMPLOYMENT MEASURED?**

Measuring unemployment is the job of the Bureau of Labor Statistics (BLS), which is part of the Department of Labor. Every month, the BLS produces data on unemployment and on other aspects of the labor market, including types of employment, length of the average workweek, and the duration of unemployment. These data come from a regular survey of about 60,000 households, called the Current Population Survey.

Based on the answers to survey questions, the BLS places each adult (age 16 and older) of each surveyed household into one of three categories:

- **Employed:** This category includes those who worked as paid employees, worked in their own business, or worked as unpaid workers in a family member’s business. Both full-time and part-time workers are counted. This category also includes those who were not working but who had jobs from which they were temporarily absent because of, for example, vacation, illness, or bad weather.
- **Unemployed:** This category includes those who were not employed, were available for work, and had tried to find employment during the previous 4 weeks. It also includes those waiting to be recalled to a job from which they had been laid off.
• **Not in the labor force:** This category includes those who fit neither of the first two categories, such as a full-time student, homemaker, or retiree.

Figure 1 shows the breakdown into these categories for 2007.

Once the BLS has placed all the individuals covered by the survey in a category, it computes various statistics to summarize the state of the labor market. The BLS defines the **labor force** as the sum of the employed and the unemployed:

\[
\text{Labor force} = \text{Number of employed} + \text{Number of unemployed}. \]

The BLS defines the **unemployment rate** as the percentage of the labor force that is unemployed:

\[
\text{Unemployment rate} = \frac{\text{Number of unemployed}}{\text{Labor force}} \times 100.
\]

The BLS computes unemployment rates for the entire adult population and for more narrowly defined groups such as blacks, whites, men, women, and so on.

The BLS uses the same survey to produce data on labor-force participation. The **labor-force participation rate** measures the percentage of the total adult population of the United States that is in the labor force:

\[
\text{Labor-force participation rate} = \frac{\text{Labor force}}{\text{Adult population}} \times 100.
\]

---

**FIGURE 1**

The Breakdown of the Population in 2007

The Bureau of Labor Statistics divides the adult population into three categories: employed, unemployed, and not in the labor force.

**Source:** Bureau of Labor Statistics.
This statistic tells us the fraction of the population that has chosen to participate in the labor market. The labor-force participation rate, like the unemployment rate, is computed for both the entire adult population and more specific groups.

To see how these data are computed, consider the figures for 2007. In that year, 146.0 million people were employed, and 7.1 million people were unemployed. The labor force was

\[ \text{Labor force} = 146.0 + 7.1 = 153.1 \text{ million}. \]

The unemployment rate was

\[ \text{Unemployment rate} = \left(\frac{7.1}{153.1}\right) \times 100 = 4.6 \text{ percent}. \]

Because the adult population was 231.8 million, the labor-force participation rate was

\[ \text{Labor-force participation rate} = \left(\frac{153.1}{231.8}\right) \times 100 = 66.0 \text{ percent}. \]

Hence, in 2007, two-thirds of the U.S. adult population were participating in the labor market, and 4.6 percent of those labor-market participants were without work.

Table 1 shows the statistics on unemployment and labor-force participation for various groups within the U.S. population. Three comparisons are most apparent. First, women ages 20 and older have lower rates of labor-force participation than men, but once in the labor force, men and women have similar rates of unemployment. Second, blacks ages 20 and older have similar rates of labor-force participation as whites, but they have much higher rates of unemployment. Third, teenagers have lower rates of labor-force participation and much higher rates of unemployment than older workers. More generally, these data show that labor-market experiences vary widely among groups within the economy.

The BLS data on the labor market also allow economists and policymakers to monitor changes in the economy over time. Figure 2 shows the unemployment

<table>
<thead>
<tr>
<th>Demographic Group</th>
<th>Unemployment Rate</th>
<th>Labor-Force Participation Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults (ages 20 and older)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, male</td>
<td>3.7%</td>
<td>76.3%</td>
</tr>
<tr>
<td>White, female</td>
<td>3.6</td>
<td>60.1</td>
</tr>
<tr>
<td>Black, male</td>
<td>7.9</td>
<td>71.2</td>
</tr>
<tr>
<td>Black, female</td>
<td>6.7</td>
<td>64.0</td>
</tr>
<tr>
<td>Teenagers (ages 16–19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, male</td>
<td>15.7</td>
<td>44.3</td>
</tr>
<tr>
<td>White, female</td>
<td>12.1</td>
<td>44.6</td>
</tr>
<tr>
<td>Black, male</td>
<td>33.8</td>
<td>29.4</td>
</tr>
<tr>
<td>Black, female</td>
<td>25.3</td>
<td>31.2</td>
</tr>
</tbody>
</table>

This table shows the unemployment rate and the labor-force participation rate of various groups in the U.S. population for 2007.

rate in the United States since 1960. The figure shows that the economy always has some unemployment and that the amount changes from year to year. The normal rate of unemployment around which the unemployment rate fluctuates is called the natural rate of unemployment, and the deviation of unemployment from its natural rate is called cyclical unemployment. The natural rate of unemployment shown in the figure is a series estimated by economists at the Congressional Budget Office. For 2007, they estimated a natural rate of 4.8 percent, close to the actual unemployment rate of 4.6 percent. Later in this book, we discuss short-run economic fluctuations, including the year-to-year fluctuations in unemployment around its natural rate. In the rest of this chapter, however, we ignore the short-run fluctuations and examine why there is always some unemployment in market economies.

LABOR-FORCE PARTICIPATION OF MEN AND WOMEN IN THE U.S. ECONOMY

Women’s role in American society has changed dramatically over the past century. Social commentators have pointed to many causes for this change. In part, it is attributable to new technologies, such as the washing machine, clothes dryer, refrigerator, freezer, and dishwasher, which have reduced the amount of time required to complete routine household tasks. In part, it is attributable to improved birth control, which has reduced the number of children born to the typical family. This change in women’s role is also partly attributable to changing political
and social attitudes, which in turn may have been facilitated by the advances in technology and birth control. Together these developments have had a profound impact on society in general and on the economy in particular.

Nowhere is that impact more obvious than in data on labor-force participation. Figure 3 shows the labor-force participation rates of men and women in the United States since 1950. Just after World War II, men and women had very different roles in society. Only 33 percent of women were working or looking for work, in contrast to 87 percent of men. Over the past several decades, the difference between the participation rates of men and women has gradually diminished, as growing numbers of women have entered the labor force and some men have left it. Data for 2007 show that 59 percent of women were in the labor force, in contrast to 73 percent of men. As measured by labor-force participation, men and women are now playing a more equal role in the economy.

The increase in women’s labor-force participation is easy to understand, but the fall in men’s may seem puzzling. There are several reasons for this decline. First,
young men now stay in school longer than their fathers and grandfathers did. Second, older men now retire earlier and live longer. Third, with more women employed, more fathers now stay at home to raise their children. Full-time students, retirees, and stay-at-home dads are all counted as being out of the labor force.

**Does the Unemployment Rate Measure What We Want It To?**

Measuring the amount of unemployment in the economy might seem a straightforward task, but it is not. While it is easy to distinguish between a person with a full-time job and a person who is not working at all, it is much harder to distinguish between a person who is unemployed and a person who is not in the labor force.
Movements into and out of the labor force are, in fact, common. More than one-
third of the unemployed are recent entrants into the labor force. These entrants
include young workers looking for their first jobs, such as recent college gradu-
ates. They also include, in greater numbers, older workers who had previously
left the labor force but have now returned to look for work. Moreover, not all
unemployment ends with the job seeker finding a job. Almost half of all spells of
unemployment end when the unemployed person leaves the labor force.

Because people move into and out of the labor force so often, statistics on
unemployment are difficult to interpret. On the one hand, some of those who
report being unemployed may not, in fact, be trying hard to find a job. They may
be calling themselves unemployed because they want to qualify for a government
program that financially assists the unemployed or because they are actually
working but paid “under the table” to avoid taxes on their earnings. It may be
more realistic to view these individuals as out of the labor force or, in some cases,
employed. On the other hand, some of those who report being out of the labor
force may want to work. These individuals may have tried to find a job and may
have given up after an unsuccessful search. Such individuals, called **discouraged
workers**, do not show up in unemployment statistics, even though they are truly
workers without jobs.

Because of these and other problems, the BLS calculates several other measures
of labor underutilization, in addition to the official unemployment rate. These
alternative measures are presented in Table 2. In the end, it is best to view the official unemployment rate as a useful but imperfect measure of joblessness.

**How Long Are the Unemployed without Work?**

In judging how serious the problem of unemployment is, one question to consider is whether unemployment is typically a short-term or long-term condition. If unemployment is short-term, one might conclude that it is not a big problem. Workers may require a few weeks between jobs to find the openings that best suit their tastes and skills. Yet if unemployment is long-term, one might conclude that it is a serious problem. Workers unemployed for many months are more likely to suffer economic and psychological hardship.

Because the duration of unemployment can affect our view about how big a problem unemployment is, economists have devoted much energy to studying data on the duration of unemployment spells. In this work, they have uncovered a result that is important, subtle, and seemingly contradictory: *Most spells of unemployment are short, and most unemployment observed at any given time is long-term.*

To see how this statement can be true, consider an example. Suppose that you visited the government’s unemployment office every week for a year to survey the unemployed. Each week you find that there are four unemployed workers. Three of these workers are the same individuals for the whole year, while the fourth person changes every week. Based on this experience, would you say that unemployment is typically short-term or long-term?

### Alternative Measures of Labor Underutilization

The table shows various measures of joblessness for the U.S. economy. The data are for February 2008.

**Source:** U.S. Department of Labor.

<table>
<thead>
<tr>
<th>Measure and Description</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-1 Persons unemployed 15 weeks or longer, as a percentage of the civilian labor force (includes only very long-term unemployed)</td>
<td>1.6%</td>
</tr>
<tr>
<td>U-2 Job losers and persons who have completed temporary jobs, as a percentage of the civilian labor force (excludes job leavers)</td>
<td>2.5</td>
</tr>
<tr>
<td>U-3 Total unemployed, as a percentage of the civilian labor force (official unemployment rate)</td>
<td>4.8</td>
</tr>
<tr>
<td>U-4 Total unemployed, plus discouraged workers, as a percentage of the civilian labor force plus discouraged workers</td>
<td>5.1</td>
</tr>
<tr>
<td>U-5 Total unemployed plus all marginally attached workers, as a percentage of the civilian labor force plus all marginally attached workers</td>
<td>5.8</td>
</tr>
<tr>
<td>U-6 Total unemployed, plus all marginally attached workers, plus total employed part-time for economic reasons, as a percentage of the civilian labor force plus all marginally attached workers</td>
<td>8.9</td>
</tr>
</tbody>
</table>

*Note:* The Bureau of Labor Statistics defines terms as follows:

- **Marginally attached workers** are persons who currently are neither working nor looking for work but indicate that they want and are available for a job and have looked for work sometime in the recent past.
- **Discouraged workers** are marginally attached workers who have given a job-market-related reason for not currently looking for a job.
- **Persons employed part-time for economic reasons** are those who want and are available for full-time work but have had to settle for a part-time schedule.
Some simple calculations help answer this question. In this example, you meet a total of 55 unemployed people over the course of a year; 52 of them are unemployed for 1 week, and 3 are unemployed for the full year. This means that 52/55, or 95 percent, of unemployment spells end in 1 week. Yet whenever you walk into the unemployment office, three of the four people you meet will be unemployed for the entire year. So, even though 95 percent of unemployment spells end in 1 week, 75 percent of the unemployment observed at any moment is attributable to those individuals who are unemployed for a full year. In this example, as in the world, most spells of unemployment are short, and most unemployment observed at any given time is long-term.

This subtle conclusion implies that economists and policymakers must be careful when interpreting data on unemployment and when designing policies to help the unemployed. Most people who become unemployed will soon find jobs. Yet most of the economy’s unemployment problem is attributable to the relatively few workers who are jobless for long periods of time.

**Why Are There Always Some People Unemployed?**

We have discussed how the government measures the amount of unemployment, the problems that arise in interpreting unemployment statistics, and the findings of labor economists on the duration of unemployment. You should now have a good idea about what unemployment is.

This discussion, however, has not explained why economies experience unemployment. In most markets in the economy, prices adjust to bring quantity supplied and quantity demanded into balance. In an ideal labor market, wages would adjust to balance the quantity of labor supplied and the quantity of labor demanded. This adjustment of wages would ensure that all workers are always fully employed.

Of course, reality does not resemble this ideal. There are always some workers without jobs, even when the overall economy is doing well. In other words, the unemployment rate never falls to zero; instead, it fluctuates around the natural rate of unemployment. To understand this natural rate, the remaining sections of this chapter examine the reasons actual labor markets depart from the ideal of full employment.

To preview our conclusions, we will find that there are four ways to explain unemployment in the long run. The first explanation is that it takes time for workers to search for the jobs that are best suited for them. The unemployment that results from the process of matching workers and jobs is sometimes called **frictional unemployment**, and it is often thought to explain relatively short spells of unemployment.

The next three explanations for unemployment suggest that the number of jobs available in some labor markets may be insufficient to give a job to everyone who wants one. This occurs when the quantity of labor supplied exceeds the quantity demanded. Unemployment of this sort is sometimes called **structural unemployment**, and it is often thought to explain longer spells of unemployment. As we will see, this kind of unemployment results when wages are, for some reason, set above the level that brings supply and demand into equilibrium. We will examine three possible reasons for an above-equilibrium wage: minimum-wage laws, unions, and efficiency wages.
QUICK QUIZ How is the unemployment rate measured? • How might the unemployment rate overstate the amount of joblessness? How might it understate the amount of joblessness?

JOB SEARCH

One reason economies always experience some unemployment is job search. Job search is the process of matching workers with appropriate jobs. If all workers and all jobs were the same, so that all workers were equally well suited for all jobs, job search would not be a problem. Laid-off workers would quickly find new jobs that were well suited for them. But in fact, workers differ in their tastes and skills, jobs differ in their attributes, and information about job candidates and job vacancies is disseminated slowly among the many firms and households in the economy.

WHY SOME FRICTIONAL UNEMPLOYMENT IS INEVITABLE

Frictional unemployment is often the result of changes in the demand for labor among different firms. When consumers decide that they prefer Dell to Apple computers, Dell increases employment, and Apple lays off workers. The former
Apple workers must now search for new jobs, and Dell must decide which new workers to hire for the various jobs that have opened up. The result of this transition is a period of unemployment.

Similarly, because different regions of the country produce different goods, employment can rise in one region while it falls in another. Consider, for instance, what happens when the world price of oil falls. Oil-producing firms in Texas respond to the lower price by cutting back on production and employment. At the same time, cheaper gasoline stimulates car sales, so auto-producing firms in Michigan raise production and employment. Just the opposite happens when the world price of oil rises. Changes in the composition of demand among industries or regions are called sectoral shifts. Because it takes time for workers to search for jobs in the new sectors, sectoral shifts temporarily cause unemployment.

Frictional unemployment is inevitable simply because the economy is always changing. A century ago, the four industries with the largest employment in the United States were cotton goods, woolen goods, men’s clothing, and lumber. Today, the four largest industries are autos, aircraft, communications, and electrical components. As this transition took place, jobs were created in some firms and destroyed in others. The result of this process has been higher productivity and higher living standards. But along the way, workers in declining industries found themselves out of work and searching for new jobs.

Data show that at least 10 percent of U.S. manufacturing jobs are destroyed every year. In addition, more than 3 percent of workers leave their jobs in a typical month, sometimes because they realize that the jobs are not a good match for their tastes and skills. Many of these workers, especially younger ones, find new jobs at higher wages. This churning of the labor force is normal in a well-functioning and dynamic market economy, but the result is some amount of frictional unemployment.

**Public Policy and Job Search**

Even if some frictional unemployment is inevitable, the precise amount is not. The faster information spreads about job openings and worker availability, the more rapidly the economy can match workers and firms. The Internet, for instance, may help facilitate job search and reduce frictional unemployment. In addition, public policy may play a role. If policy can reduce the time it takes unemployed workers to find new jobs, it can reduce the economy’s natural rate of unemployment.

Government programs try to facilitate job search in various ways. One way is through government-run employment agencies, which give out information about job vacancies. Another way is through public training programs, which aim to ease the transition of workers from declining to growing industries and to help disadvantaged groups escape poverty. Advocates of these programs believe that they make the economy operate more efficiently by keeping the labor force more fully employed and that they reduce the inequities inherent in a constantly changing market economy.

Critics of these programs question whether the government should get involved with the process of job search. They argue that it is better to let the private market match workers and jobs. In fact, most job searching in our economy takes place without intervention by the government. Newspaper ads, Internet job sites, college placement offices, headhunters, and word of mouth all help spread informa-
tion about job openings and job candidates. Similarly, much worker education is done privately, either through schools or through on-the-job training. These critics contend that the government is no better—and most likely worse—at disseminating the right information to the right workers and deciding what kinds of worker training would be most valuable. They claim that these decisions are best made privately by workers and employers.

Unemployment Insurance

One government program that increases the amount of frictional unemployment, without intending to do so, is unemployment insurance. This program is designed to offer workers partial protection against job loss. The unemployed who quit their jobs, were fired for cause, or just entered the labor force are not eligible. Benefits are paid only to the unemployed who were laid off because their previous employers no longer needed their skills. The terms of the program vary over time and across states, but a typical worker covered by unemployment insurance in the United States receives 50 percent of his or her former wages for 26 weeks.

While unemployment insurance reduces the hardship of unemployment, it also increases the amount of unemployment. The explanation is based on one of the Ten Principles of Economics in Chapter 1: People respond to incentives. Because unemployment benefits stop when a worker takes a new job, the unemployed devote less effort to job search and are more likely to turn down unattractive job offers. In addition, because unemployment insurance makes unemployment less onerous, workers are less likely to seek guarantees of job security when they negotiate with employers over the terms of employment.

Many studies by labor economists have examined the incentive effects of unemployment insurance. One study examined an experiment run by the state of Illinois in 1985. When unemployed workers applied to collect unemployment insurance benefits, the state randomly selected some of them and offered each a $500 bonus if they found new jobs within 11 weeks. This group was then compared to a control group not offered the incentive. The average spell of unemployment for the group offered the bonus was 7 percent shorter than the average spell for the control group. This experiment shows that the design of the unemployment insurance system influences the effort that the unemployed devote to job search.

Several other studies examined search effort by following a group of workers over time. Unemployment insurance benefits, rather than lasting forever, usually run out after 6 months or 1 year. These studies found that when the unemployed become ineligible for benefits, the probability of their finding a new job rises markedly. Thus, receiving unemployment insurance benefits does reduce the search effort of the unemployed.

Even though unemployment insurance reduces search effort and raises unemployment, we should not necessarily conclude that the policy is bad. The program does achieve its primary goal of reducing the income uncertainty that workers face. In addition, when workers turn down unattractive job offers, they have the opportunity to look for jobs that better suit their tastes and skills. Some economists argue that unemployment insurance improves the ability of the economy to match each worker with the most appropriate job.

The study of unemployment insurance shows that the unemployment rate is an imperfect measure of a nation’s overall level of economic well-being. Most
Traditionally, many European countries have had unemployment insurance that is far more generous than that offered to U.S. workers. But policymakers on both continents are starting to reconsider.

Faced with relentless competition from workers in China and other poor countries, American policy makers are considering extending a little more generosity to the low-skilled and unemployed. In Europe, the debate focuses on pushing the unemployed back to work.

"There is a convergence across the Atlantic," said Hans-Werner Sinn, president of the Ifo Institute for Economic Research in Munich, whose proposals to get the unemployed back to work have been partially adopted in Germany.

In the United States, after Democrats took hold of Congress, the Senate and the House passed separate bills to raise the minimum wage. If the legislation is enacted, it will be the first minimum wage increase in a decade. [Author’s update: An increase in the minimum wage passed and was signed into law in May 2007.]

Now Democratic lawmakers are considering a plan to extend unemployment insurance coverage, which excludes all sorts of laid-off workers, including those who had part-time jobs and those leaving work because of family needs. Some Democrats are also proposing a wage insurance plan, intended to cover part of the difference between the wages a laid-off worker formerly earned and the wages paid in the new job, which are typically 16 percent lower.

“You can think of it as subsidizing the re-acquisition of skills by workers,” said Lael Brainard, an economist at the Brookings Institution in Washington.

Representative Jim McDermott, a Democrat from Washington State who has drafted proposals to retool unemployment insurance and introduce wage insurance, said: "In the mid-1990s, globalization accelerated to 100 miles per hour. Unfortunately nobody was standing around and thinking about what was happening to American workers."

Most economists argue that American workers have long been buffeted by similar forces. "I wouldn't think of globalization as fundamentally different from other changes the United States economy has undergone over hundreds of years," said Robert Shimer, a professor of economics at the University of Chicago. "Now jobs are lost in part because they are going to other countries, but buggy makers lost their jobs when the car was introduced."

Economists agree that eliminating unemployment insurance would reduce the amount of unemployment in the economy. Yet economists disagree on whether economic well-being would be enhanced or diminished by this change in policy.

**QUICK QUIZ** How would an increase in the world price of oil affect the amount of frictional unemployment? Is this unemployment undesirable? What public policies might affect the amount of unemployment caused by this price change?
There is an ideological component in the policy proposals to address the current dislocations. In an address to the American Economic Association two years ago, Martin Feldstein, a former economic adviser to President Bush, suggested that unemployment insurance could be replaced with private unemployment accounts—similar in approach to private retirement accounts—for workers to prepare for potential joblessness.

Critics of the wage insurance proposal argue that it would subsidize employers who pay substandard wages. In addition, some say that it is inefficient to distort workers’ incentives this way. Some Republican-leaning economists support the new ideas, in principle. “There’s a clear social contract between capital and labor,” said John Silvia, chief economist of the Wachovia Corporation—who was formerly an economist for two Republican-controlled Congressional committees. “When a corporation reallocates its resources we cannot just leave these people flat and dead.”

Moreover, many economists say, helping workers deal with the dislocations wrought by a fast-changing economy would help temper their fears about globalization and forestall a protectionist backlash.

Global competition from manufacturers in poor countries, meanwhile, is driving some Western European countries in the opposite direction. Realizing that high long-term unemployment benefits established a very high “reservation wage”—below which a person who is out of work would not accept a job—many are trying to coax the unemployed back into active work.

In Denmark, home to one of Europe’s most generous unemployment programs, public munificence has eroded somewhat in the last decade. The unemployed must now accept either a job offer or a place in a training program if they are to keep their benefits; the maximum benefit period has been cut to four years from five.

In Germany, the former left-of-center government passed a package of laws in 2003 that cut benefits, pared the duration of unemployment insurance to 16 months from 32 and required workers on long-term benefits to accept any “reasonable” job offers. . . .

Europeans and Americans take very different approaches to the labor market—with Europeans concerned primarily with guaranteeing some decent level of income for its citizens, the United States with encouraging an efficient labor market that provides lots of jobs. Still, a common perspective may be emerging: the unemployed must be encouraged to work. But it may be the government’s role to step in when wages are insufficient.

<table>
<thead>
<tr>
<th>Country</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>8%</td>
</tr>
<tr>
<td>Greece</td>
<td>13%</td>
</tr>
<tr>
<td>United States</td>
<td>14%</td>
</tr>
<tr>
<td>Canada</td>
<td>15%</td>
</tr>
<tr>
<td>Britain</td>
<td>16%</td>
</tr>
<tr>
<td>Germany</td>
<td>29%</td>
</tr>
<tr>
<td>Switzerland</td>
<td>33%</td>
</tr>
<tr>
<td>Italy</td>
<td>34%</td>
</tr>
<tr>
<td>Finland</td>
<td>39%</td>
</tr>
<tr>
<td>France</td>
<td>50%</td>
</tr>
<tr>
<td>Denmark</td>
<td>53%</td>
</tr>
</tbody>
</table>


**MINIMUM-WAGE LAWS**

Having seen how frictional unemployment results from the process of matching workers and jobs, let’s now examine how structural unemployment results when the number of jobs is insufficient for the number of workers.

To understand structural unemployment, we begin by reviewing how minimum-wage laws can cause unemployment. Although minimum wages are
not the predominant reason for unemployment in our economy, they have an important effect on certain groups with particularly high unemployment rates. Moreover, the analysis of minimum wages is a natural place to start because, as we will see, it can be used to understand some of the other reasons for structural unemployment.

Figure 4 reviews the basic economics of a minimum wage. When a minimum-wage law forces the wage to remain above the level that balances supply and demand, it raises the quantity of labor supplied and reduces the quantity of labor demanded compared to the equilibrium level. There is a surplus of labor. Because there are more workers willing to work than there are jobs, some workers are unemployed.

While minimum-wage laws are one reason unemployment exists in the U.S. economy, they do not affect everyone. Most workers have wages well above the legal minimum, so the law does not prevent the wage from adjusting to balance supply and demand. Minimum-wage laws matter most for the least skilled and least experienced members of the labor force, such as teenagers. Their equilibrium wages tend to be low and, therefore, are more likely to fall below the legal minimum. It is only among these workers that minimum-wage laws explain the existence of unemployment.

Figure 4 is drawn to show the effects of a minimum-wage law, but it also illustrates a more general lesson: If the wage is kept above the equilibrium level for any reason, the result is unemployment. Minimum-wage laws are just one reason wages may be “too high.” In the remaining two sections of this chapter, we consider two other reasons wages may be kept above the equilibrium level: unions and efficiency wages. The basic economics of unemployment in these cases is the same as that shown in Figure 4, but these explanations of unemployment can apply to many more of the economy’s workers.
At this point, however, we should stop and notice that the structural unemployment that arises from an above-equilibrium wage is, in an important sense, different from the frictional unemployment that arises from the process of job search. The need for job search is not due to the failure of wages to balance labor supply and labor demand. When job search is the explanation for unemployment, workers are searching for the jobs that best suit their tastes and skills. By contrast, when the wage is above the equilibrium level, the quantity of labor supplied exceeds the quantity of labor demanded, and workers are unemployed because they are waiting for jobs to open up.

**Quick Quiz** Draw the supply curve and the demand curve for a labor market in which the wage is fixed above the equilibrium level. Show the quantity of labor supplied, the quantity demanded, and the amount of unemployment.

**UNIONS AND COLLECTIVE BARGAINING**

A union is a worker association that bargains with employers over wages, benefits, and working conditions. Whereas only 12 percent of U.S. workers now belong to unions, unions played a much larger role in the U.S. labor market in the past.
In the 1940s and 1950s, when unions were at their peak, about a third of the U.S. labor force was unionized.

Moreover, for a variety of historical reasons, unions continue to play a large role in many European countries. In Belgium, Norway, and Sweden, for instance, more than half of workers belong to unions. In France and Germany, a majority of workers have wages set by collective bargaining by law, even though only some of these workers are themselves union members. In these cases, wages are not determined by the equilibrium of supply and demand in competitive labor markets.

**The Economics of Unions**

A union is a type of cartel. Like any cartel, a union is a group of sellers acting together in the hope of exerting their joint market power. Most workers in the U.S. economy discuss their wages, benefits, and working conditions with their employers as individuals. By contrast, workers in a union do so as a group. The process by which unions and firms agree on the terms of employment is called collective bargaining.

When a union bargains with a firm, it asks for higher wages, better benefits, and better working conditions than the firm would offer in the absence of a union. If the union and the firm do not reach agreement, the union can organize a withdrawal of labor from the firm, called a strike. Because a strike reduces production, sales, and profit, a firm facing a strike threat is likely to agree to pay higher wages than it otherwise would. Economists who study the effects of unions typically find that union workers earn about 10 to 20 percent more than similar workers who do not belong to unions.

When a union raises the wage above the equilibrium level, it raises the quantity of labor supplied and reduces the quantity of labor demanded, resulting in unemployment. Workers who remain employed at the higher wage are better off, but those who were previously employed and are now unemployed are worse off. Indeed, unions are often thought to cause conflict between different groups of workers—between the insid-

ers who benefit from high union wages and the outsiders who do not get the union jobs.

The outsiders can respond to their status in one of two ways. Some of them remain unemployed and wait for the chance to become insiders and earn the high union wage. Others take jobs in firms that are not unionized. Thus, when unions raise wages in one part of the economy, the supply of labor increases in other parts of the economy. This increase in labor supply, in turn, reduces wages in industries that are not unionized. In other words, workers in unions reap the benefit of collective bargaining, while workers not in unions bear some of the cost.

The role of unions in the economy depends in part on the laws that govern union organization and collective bargaining. Normally, explicit agreements among members of a cartel are illegal. When firms selling similar products agree to set high prices, the agreement is considered a “conspiracy in restraint of trade,” and the government prosecutes the firms in civil and criminal court for violating the antitrust laws. By contrast, unions are exempt from these laws. The policymakers who wrote the antitrust laws believed that workers needed greater market power as they bargained with employers. Indeed, various laws are designed to encourage the formation of unions. In particular, the Wagner Act of 1935 prevents
employers from interfering when workers try to organize unions and requires employers to bargain with unions in good faith. The National Labor Relations Board (NLRB) is the government agency that enforces workers’ right to unionize.

Legislation affecting the market power of unions is a perennial topic of political debate. State lawmakers sometimes debate right-to-work laws, which give workers in a unionized firm the right to choose whether to join the union. In the absence of such laws, unions can insist during collective bargaining that firms make union membership a requirement for employment. At times, lawmakers in Washington have debated a proposed law that would prevent firms from hiring permanent replacements for workers who are on strike. This law would make strikes more costly for firms, thereby increasing the market power of unions. These and similar policy decisions will help determine the future of the union movement.

**Are Unions Good or Bad for the Economy?**

Economists disagree about whether unions are good or bad for the economy as a whole. Let’s consider both sides of the debate.

Critics argue that unions are merely a type of cartel. When unions raise wages above the level that would prevail in competitive markets, they reduce the quantity of labor demanded, cause some workers to be unemployed, and reduce the wages in the rest of the economy. The resulting allocation of labor is, critics argue, both inefficient and inequitable. It is inefficient because high union wages reduce employment in unionized firms below the efficient, competitive level. It is inequitable because some workers benefit at the expense of other workers.

Advocates contend that unions are a necessary antidote to the market power of the firms that hire workers. The extreme case of this market power is the “company town,” where a single firm does most of the hiring in a geographical region. In a company town, if workers do not accept the wages and working conditions that the firm offers, they have little choice but to move or stop working. In the absence of a union, therefore, the firm could use its market power to pay lower wages and offer worse working conditions than would prevail if it had to compete with other firms for the same workers. In this case, a union may balance the firm’s market power and protect the workers from being at the mercy of the firm’s owners.

Advocates of unions also claim that unions are important for helping firms respond efficiently to workers’ concerns. Whenever a worker takes a job, the worker and the firm must agree on many attributes of the job in addition to the wage: hours of work, overtime, vacations, sick leave, health benefits, promotion schedules, job security, and so on. By representing workers’ views on these issues, unions allow firms to provide the right mix of job attributes. Even if unions have the adverse effect of pushing wages above the equilibrium level and causing unemployment, they have the benefit of helping firms keep a happy and productive workforce.

In the end, there is no consensus among economists about whether unions are good or bad for the economy. Like many institutions, their influence is probably beneficial in some circumstances and adverse in others.

**Quick Quiz**  How does a union in the auto industry affect wages and employment at General Motors and Ford? How does it affect wages and employment in other industries?
A fourth reason economies always experience some unemployment—in addition to job search, minimum-wage laws, and unions—is suggested by the theory of **efficiency wages**. According to this theory, firms operate more efficiently if wages are above the equilibrium level. Therefore, it may be profitable for firms to keep wages high even in the presence of a surplus of labor.

In some ways, the unemployment that arises from efficiency wages is similar to the unemployment that arises from minimum-wage laws and unions. In all three cases, unemployment is the result of wages above the level that balances the quantity of labor supplied and the quantity of labor demanded. Yet there is also an important difference. Minimum-wage laws and unions prevent firms from lowering wages in the presence of a surplus of workers. Efficiency-wage theory states that such a constraint on firms is unnecessary in many cases because firms may be better off keeping wages above the equilibrium level.

Why should firms want to keep wages high? This decision may seem odd at first, for wages are a large part of firms’ costs. Normally, we expect profit-maximizing firms to want to keep costs—and therefore wages—as low as possible. The novel insight of efficiency-wage theory is that paying high wages might be profitable because they might raise the efficiency of a firm’s workers.

There are several types of efficiency-wage theory. Each type suggests a different explanation for why firms may want to pay high wages. Let’s now consider four of these types.

**Worker Health**

The first and simplest type of efficiency-wage theory emphasizes the link between wages and worker health. Better paid workers eat a more nutritious diet, and workers who eat a better diet are healthier and more productive. A firm may find it more profitable to pay high wages and have healthy, productive workers than to pay lower wages and have less healthy, less productive workers.

This type of efficiency-wage theory can be relevant for explaining unemployment in less developed countries where inadequate nutrition can be a problem. In these countries, firms may fear that cutting wages would, in fact, adversely influence their workers’ health and productivity. In other words, nutrition concerns may explain why firms may maintain above-equilibrium wages despite a surplus of labor. Worker health concerns are far less relevant for firms in rich countries such as the United States, where the equilibrium wages for most workers are well above the level needed for an adequate diet.

**Worker Turnover**

A second type of efficiency-wage theory emphasizes the link between wages and worker turnover. Workers quit jobs for many reasons: to take jobs in other firms, to move to other parts of the country, to leave the labor force, and so on. The frequency with which they quit depends on the entire set of incentives they face, including the benefits of leaving and the benefits of staying. The more a firm pays its workers, the less often its workers will choose to leave. Thus, a firm can reduce turnover among its workers by paying them a high wage.
Why do firms care about turnover? The reason is that it is costly for firms to hire and train new workers. Moreover, even after they are trained, newly hired workers are not as productive as experienced workers. Firms with higher turnover, therefore, will tend to have higher production costs. Firms may find it profitable to pay wages above the equilibrium level to reduce worker turnover.

**Worker Quality**

A third type of efficiency-wage theory emphasizes the link between wages and worker quality. All firms want workers who are talented, and they try to pick the best applicants to fill job openings. But because firms cannot perfectly gauge the quality of applicants, hiring has a degree of randomness to it. When a firm pays a high wage, it attracts a better pool of workers to apply for its jobs and thereby increases the quality of its workforce. If the firm responded to a surplus of labor by reducing the wage, the most competent applicants—who are more likely to have better alternative opportunities than less competent applicants—may choose not to apply. If this influence of the wage on worker quality is strong enough, it may be profitable for the firm to pay a wage above the level that balances supply and demand.

**Worker Effort**

A fourth and final type of efficiency-wage theory emphasizes the link between wages and worker effort. In many jobs, workers have some discretion over how hard to work. As a result, firms monitor the efforts of their workers, and workers caught shirking their responsibilities are fired. But not all shirkers are caught immediately because monitoring workers is costly and imperfect. A firm in such a circumstance is always looking for ways to deter shirking.

One solution is paying wages above the equilibrium level. High wages make workers more eager to keep their jobs and, thereby, give workers an incentive to put forward their best effort. If the wage were at the level that balanced supply and demand, workers would have less reason to work hard because if they were fired, they could quickly find new jobs at the same wage. Therefore, firms raise wages above the equilibrium level, providing an incentive for workers not to shirk their responsibilities.
HENRY FORD AND THE VERY GENEROUS $5-A-DAY WAGE

Henry Ford was an industrial visionary. As founder of the Ford Motor Company, he was responsible for introducing modern techniques of production. Rather than building cars with small teams of skilled craftsmen, Ford built cars on assembly lines in which unskilled workers were taught to perform the same simple tasks over and over again. The output of this assembly process was the Model T Ford, one of the most famous early automobiles.

In 1914, Ford introduced another innovation: the $5 workday. This might not seem like much today, but back then $5 was about twice the going wage. It was also far above the wage that balanced supply and demand. When the new $5-a-day wage was announced, long lines of job seekers formed outside the Ford factories. The number of workers willing to work at this wage far exceeded the number of workers Ford needed.

Ford’s high-wage policy had many of the effects predicted by efficiency-wage theory. Turnover fell, absenteeism fell, and productivity rose. Workers were so much more efficient that Ford’s production costs were lower despite higher wages. Thus, paying a wage above the equilibrium level was profitable for the firm. An historian of the early Ford Motor Company wrote, “Ford and his associates freely declared on many occasions that the high-wage policy turned out to be good business. By this they meant that it had improved the discipline of the workers, given them a more loyal interest in the institution, and raised their personal efficiency.” Henry Ford himself called the $5-a-day wage “one of the finest cost-cutting moves we ever made.”

Why did it take Henry Ford to introduce this efficiency wage? Why were other firms not already taking advantage of this seemingly profitable business strategy? According to some analysts, Ford’s decision was closely linked to his use of the assembly line. Workers organized in an assembly line are highly interdependent. If one worker is absent or works slowly, other workers are less able to complete their own tasks. Thus, while assembly lines made production more efficient, they also raised the importance of low worker turnover, high worker effort, and high worker quality. As a result, paying efficiency wages may have been a better strategy for the Ford Motor Company than for other businesses at the time.

QUICK QUIZ Give four explanations for why firms might find it profitable to pay wages above the level that balances quantity of labor supplied and quantity of labor demanded.

CONCLUSION

In this chapter, we discussed the measurement of unemployment and the reasons economies always experience some degree of unemployment. We have seen how job search, minimum-wage laws, unions, and efficiency wages can all help explain why some workers do not have jobs. Which of these four explanations for the natural rate of unemployment are the most important for the U.S. economy and other economies around the world? Unfortunately, there is no easy way to tell. Economists differ in which of these explanations of unemployment they consider most important.
The analysis of this chapter yields an important lesson: Although the economy will always have some unemployment, its natural rate does change over time. Many events and policies can alter the amount of unemployment the economy typically experiences. As the information revolution changes the process of job search, as Congress adjusts the minimum wage, as workers form or quit unions, and as firms change their reliance on efficiency wages, the natural rate of unemployment evolves. Unemployment is not a simple problem with a simple solution. But how we choose to organize our society can profoundly influence how prevalent a problem it is.

**KEY CONCEPTS**

- The unemployment rate is the percentage of those who would like to work who do not have jobs. The Bureau of Labor Statistics calculates this statistic monthly based on a survey of thousands of households.
- The unemployment rate is an imperfect measure of joblessness. Some people who call themselves unemployed may actually not want to work, and some people who would like to work have left the labor force after an unsuccessful search and therefore are not counted as unemployed.
- In the U.S. economy, most people who become unemployed find work within a short period of time. Nonetheless, most unemployment observed at any given time is attributable to the few people who are unemployed for long periods of time.
- One reason for unemployment is the time it takes workers to search for jobs that best suit their tastes and skills. This frictional unemployment is increased as a result of unemployment insurance, a government policy designed to protect workers’ incomes.
- A second reason our economy always has some unemployment is minimum-wage laws. By raising the wage of unskilled and inexperienced workers above the equilibrium level, minimum-wage laws raise the quantity of labor supplied and reduce the quantity demanded. The resulting surplus of labor represents unemployment.
- A third reason for unemployment is the market power of unions. When unions push the wages in unionized industries above the equilibrium level, they create a surplus of labor.
- A fourth reason for unemployment is suggested by the theory of efficiency wages. According to this theory, firms find it profitable to pay wages above the equilibrium level. High wages can improve worker health, lower worker turnover, raise worker quality, and increase worker effort.

**SUMMARY**

- The unemployment rate is the percentage of those who would like to work who do not have jobs. The Bureau of Labor Statistics calculates this statistic monthly based on a survey of thousands of households.
- The unemployment rate is an imperfect measure of joblessness. Some people who call themselves unemployed may actually not want to work, and some people who would like to work have left the labor force after an unsuccessful search and therefore are not counted as unemployed.
- In the U.S. economy, most people who become unemployed find work within a short period of time. Nonetheless, most unemployment observed at any given time is attributable to the few people who are unemployed for long periods of time.
- One reason for unemployment is the time it takes workers to search for jobs that best suit their tastes and skills. This frictional unemployment is increased as a result of unemployment insurance, a government policy designed to protect workers’ incomes.
- A second reason our economy always has some unemployment is minimum-wage laws. By raising the wage of unskilled and inexperienced workers above the equilibrium level, minimum-wage laws raise the quantity of labor supplied and reduce the quantity demanded. The resulting surplus of labor represents unemployment.
- A third reason for unemployment is the market power of unions. When unions push the wages in unionized industries above the equilibrium level, they create a surplus of labor.
- A fourth reason for unemployment is suggested by the theory of efficiency wages. According to this theory, firms find it profitable to pay wages above the equilibrium level. High wages can improve worker health, lower worker turnover, raise worker quality, and increase worker effort.

**KEY CONCEPTS**

- labor force, p. 431
- unemployment rate, p. 431
- labor-force participation rate, p. 431
- natural rate of unemployment, p. 433
- cyclical unemployment, p. 433
- discouraged workers, p. 436
- frictional unemployment, p. 438
- structural unemployment, p. 438
- job search, p. 439
- unemployment insurance, p. 441
- union, p. 445
- collective bargaining, p. 446
- strike, p. 446
- efficiency wages, p. 448
**QUESTIONS FOR REVIEW**

1. What are the three categories into which the Bureau of Labor Statistics divides everyone? How does the BLS compute the labor force, the unemployment rate, and the labor-force participation rate?
2. Is unemployment typically short-term or long-term? Explain.
3. Why is frictional unemployment inevitable? How might the government reduce the amount of frictional unemployment?
4. Are minimum-wage laws a better explanation for structural unemployment among teenagers or among college graduates? Why?
5. How do unions affect the natural rate of unemployment?
6. What claims do advocates of unions make to argue that unions are good for the economy?
7. Explain four ways in which a firm might increase its profits by raising the wages it pays.

**PROBLEMS AND APPLICATIONS**

1. The Bureau of Labor Statistics announced that in February 2008, of all adult Americans, 145,993,000 were employed, 7,381,000 were unemployed, and 79,436,000 were not in the labor force. Use this information to calculate:
   a. the adult population
   b. the labor force
   c. the labor-force participation rate
   d. the unemployment rate

2. Go to the website of the Bureau of Labor Statistics (http://www.bls.gov). What is the national unemployment rate right now? Find the unemployment rate for the demographic group that best fits a description of you (for example, based on age, sex, and race). Is it higher or lower than the national average? Why do you think this is so?

3. As shown in Figure 3, the overall labor-force participation rate of men declined between 1970 and 2000. At the same time, the labor-force participation rate of women increased sharply. This overall decline reflects different patterns for different age groups, however, as shown in the following tables.

<table>
<thead>
<tr>
<th></th>
<th>All Men</th>
<th>Men 16–24</th>
<th>Men 25–54</th>
<th>Men 55 and older</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>80%</td>
<td>69%</td>
<td>96%</td>
<td>56%</td>
</tr>
<tr>
<td>2000</td>
<td>75%</td>
<td>69%</td>
<td>92%</td>
<td>40%</td>
</tr>
</tbody>
</table>

Given this information, what factor do you think played the key role in the decline in male labor-force participation over this period? What do you think explains the increase in labor-force participation for women?

4. Between 2004 and 2007, total U.S. employment increased by 6.8 million workers, but the number of unemployed workers declined by only 1.1 million. How are these numbers consistent with each other? Why might one expect a reduction in the number of people counted as unemployed to be smaller than the increase in the number of people employed?

5. Economists use labor-market data to evaluate how well an economy is using its most valuable resource—its people. Two closely watched statistics are the unemployment rate and the employment-population ratio. Explain what happens to each of these in the following scenarios. In your opinion, which statistic is the more meaningful gauge of how well the economy is doing?
   a. An auto company goes bankrupt and lays off its workers, who immediately start looking for new jobs.
a. If workers in these two industries require similar amounts of training, and if workers at the beginning of their careers could choose which industry to train for, what would you expect to happen to the wages in these two industries? How long would this process take? Explain.

b. Suppose that one day the economy opens itself to international trade and, as a result, starts importing autos and exporting aircraft. What would happen to demand for labor in these two industries?

c. Suppose that workers in one industry cannot be quickly retrained for the other. How would these shifts in demand affect equilibrium wages both in the short run and in the long run?

d. If for some reason wages fail to adjust to the new equilibrium levels, what would occur?

10. Suppose that Congress passes a law requiring employers to provide employees some benefit (such as healthcare) that raises the cost of an employee by $4 per hour.

a. What effect does this employer mandate have on the demand for labor? (In answering this and the following questions, be quantitative when you can.)

b. If employees place a value on this benefit exactly equal to its cost, what effect does this employer mandate have on the supply of labor?

c. If the wage is free to balance supply and demand, how does this law affect the wage and the level of employment? Are employers better or worse off? Are employees better or worse off?

d. Suppose that, before the mandate, the wage in this market was $3 above the minimum wage. In this case, how does the employer mandate affect the wage, the level of employment, and the level of unemployment?

e. Now suppose that workers do not value the mandated benefit at all. How does this alternative assumption change your answers to parts (b) and (c)?

b. After an unsuccessful search, some of the laid-off workers quit looking for new jobs.

c. Numerous students graduate from college but cannot find work.

d. Numerous students graduate from college and immediately begin new jobs.

e. A stock market boom induces newly enriched 60-year-old workers to take early retirement.

f. Advances in health care prolong the life of many retirees.

6. Are the following workers more likely to experience short-term or long-term unemployment? Explain.

a. A construction worker laid off because of bad weather

b. A manufacturing worker who loses her job at a plant in an isolated area

c. A stagecoach-industry worker laid off because of competition from railroads

d. A short-order cook who loses his job when a new restaurant opens across the street

e. An expert welder with little formal education who loses her job when the company installs automatic welding machinery

7. Using a diagram of the labor market, show the effect of an increase in the minimum wage on the wage paid to workers, the number of workers supplied, the number of workers demanded, and the amount of unemployment.

8. Consider an economy with two labor markets—one for manufacturing workers and one for service workers. Suppose initially that neither is unionized.

a. If manufacturing workers formed a union, what impact on the wages and employment in manufacturing would you predict?

b. How would these changes in the manufacturing labor market affect the supply of labor in the market for service workers? What would happen to the equilibrium wage and employment in this labor market?

9. Structural unemployment is sometimes said to result from a mismatch between the job skills that employers want and the job skills that workers have. To explore this idea, consider an economy with two industries: auto manufacturing and aircraft manufacturing.
PART VIII

Money and Prices in the Long Run
This page intentionally left blank
The Monetary System

When you walk into a restaurant to buy a meal, you get something of value—a full stomach. To pay for this service, you might hand the restaurateur several worn-out pieces of greenish paper decorated with strange symbols, government buildings, and the portraits of famous dead Americans. Or you might hand him a single piece of paper with the name of a bank and your signature. Whether you pay by cash or check, the restaurateur is happy to work hard to satisfy your gastronomical desires in exchange for these pieces of paper which, in and of themselves, are worthless.

To anyone who has lived in a modern economy, this social custom is not at all odd. Even though paper money has no intrinsic value, the restaurateur is confident that, in the future, some third person will accept it in exchange for something that the restaurateur does value. And that third person is confident that some fourth person will accept the money, with the knowledge that yet a fifth person will accept the money . . . and so on. To the restaurateur and to other people in our society, your cash or check represents a claim to goods and services in the future.

The social custom of using money for transactions is extraordinarily useful in a large, complex society. Imagine, for a moment, that there was no item in the economy widely accepted in exchange for goods and services. People would have to rely on barter—the exchange of one good or service for another—to obtain the things they need. To get your restaurant meal, for instance, you would have to
offer the restaurateur something of immediate value. You could offer to wash some dishes, clean his car, or give him your family’s secret recipe for meat loaf. An economy that relies on barter will have trouble allocating its scarce resources efficiently. In such an economy, trade is said to require the double coincidence of wants—the unlikely occurrence that two people each have a good or service that the other wants.

The existence of money makes trade easier. The restaurateur does not care whether you can produce a valuable good or service for him. He is happy to accept your money, knowing that other people will do the same for him. Such a convention allows trade to be roundabout. The restaurateur accepts your money and uses it to pay his chef; the chef uses her paycheck to send her child to day care; the day care center uses this tuition to pay a teacher; and the teacher hires you to mow his lawn. As money flows from person to person in the economy, it facilitates production and trade, thereby allowing each person to specialize in what he or she does best and raising everyone’s standard of living.

In this chapter, we begin to examine the role of money in the economy. We discuss what money is, the various forms that money takes, how the banking system helps create money, and how the government controls the quantity of money in circulation. Because money is so important in the economy, we devote much effort in the rest of this book to learning how changes in the quantity of money affect various economic variables, including inflation, interest rates, production, and employment. Consistent with our long-run focus in the previous four chapters, in the next chapter we will examine the long-run effects of changes in the quantity of money. The short-run effects of monetary changes are a more complex topic, which we will take up later in the book. This chapter provides the background for all of this further analysis.

**THE MEANING OF MONEY**

What is money? This might seem like an odd question. When you read that billionaire Bill Gates has a lot of money, you know what that means: He is so rich that he can buy almost anything he wants. In this sense, the term *money* is used to mean *wealth*.

Economists, however, use the word in a more specific sense: *Money* is the set of assets in the economy that people regularly use to buy goods and services from each other. The cash in your wallet is money because you can use it to buy a meal at a restaurant or a shirt at a clothing store. By contrast, if you happened to own most of Microsoft Corporation, as Bill Gates does, you would be wealthy, but this asset is not considered a form of money. You could not buy a meal or a shirt with this wealth without first obtaining some cash. According to the economist’s definition, money includes only those few types of wealth that are regularly accepted by sellers in exchange for goods and services.

**THE FUNCTIONS OF MONEY**

Money has three functions in the economy: It is a *medium of exchange*, a *unit of account*, and a *store of value*. These three functions together distinguish money from other assets in the economy, such as stocks, bonds, real estate, art, and even baseball cards. Let’s examine each of these functions of money in turn.
A **medium of exchange** is an item that buyers give to sellers when they purchase goods and services. When you buy a shirt at a clothing store, the store gives you the shirt, and you give the store your money. This transfer of money from buyer to seller allows the transaction to take place. When you walk into a store, you are confident that the store will accept your money for the items it is selling because money is the commonly accepted medium of exchange.

A **unit of account** is the yardstick people use to post prices and record debts. When you go shopping, you might observe that a shirt costs $20 and a hamburger costs $2. Even though it would be accurate to say that the price of a shirt is 10 hamburgers and the price of a hamburger is 1/10 of a shirt, prices are never quoted in this way. Similarly, if you take out a loan from a bank, the size of your future loan repayments will be measured in dollars, not in a quantity of goods and services. When we want to measure and record economic value, we use money as the unit of account.

A **store of value** is an item that people can use to transfer purchasing power from the present to the future. When a seller accepts money today in exchange for a good or service, that seller can hold the money and become a buyer of another good or service at another time. Money is not the only store of value in the economy: A person can also transfer purchasing power from the present to the future by holding nonmonetary assets such as stocks and bonds. The term *wealth* is used to refer to the total of all stores of value, including both money and nonmonetary assets.

Economists use the term **liquidity** to describe the ease with which an asset can be converted into the economy’s medium of exchange. Because money is the economy’s medium of exchange, it is the most liquid asset available. Other assets vary widely in their liquidity. Most stocks and bonds can be sold easily with small cost, so they are relatively liquid assets. By contrast, selling a house, a Rembrandt painting, or a 1948 Joe DiMaggio baseball card requires more time and effort, so these assets are less liquid.

When people decide in what form to hold their wealth, they have to balance the liquidity of each possible asset against the asset’s usefulness as a store of value. Money is the most liquid asset, but it is far from perfect as a store of value. When prices rise, the value of money falls. In other words, when goods and services become more expensive, each dollar in your wallet can buy less. This link between the price level and the value of money is key to understanding how money affects the economy, a topic we will start to explore in the next chapter.

**The Kinds of Money**

When money takes the form of a commodity with intrinsic value, it is called **commodity money**. The term *intrinsic value* means that the item would have value even if it were not used as money. One example of commodity money is gold. Gold has intrinsic value because it is used in industry and in the making of jewelry. Although today we no longer use gold as money, historically gold has been a common form of money because it is relatively easy to carry, measure, and verify for impurities. When an economy uses gold as money (or uses paper money that is convertible into gold on demand), it is said to be operating under a gold standard.

Another example of commodity money is cigarettes. In prisoner-of-war camps during World War II, prisoners traded goods and services with one another using cigarettes as the store of value, unit of account, and medium of exchange. Similarly, as the Soviet Union was breaking up in the late 1980s, cigarettes started replacing the ruble as the preferred currency in Moscow. In both cases, even nonsmokers...
money that is used as money because of government decree

fiat money

were happy to accept cigarettes in an exchange, knowing that they could use the cigarettes to buy other goods and services.

Money without intrinsic value is called **fiat money**. A *fiat* is an order or decree, and fiat money is established as money by government decree. For example, compare the paper dollars in your wallet (printed by the U.S. government) and the paper dollars from a game of Monopoly (printed by the Parker Brothers game company). Why can you use the first to pay your bill at a restaurant but not the second? The answer is that the U.S. government has decreed its dollars to be valid money. Each paper dollar in your wallet reads: “This note is legal tender for all debts, public and private. Gold, cigarettes, cowrie shells and peacock feathers all have “use value” in addition to their “exchange value.” These items were originally valued for their utility or their beauty, and they became used as currency. It is rare to see a purely paper currency functioning as money without the backing of some government or financial institution.

Rare, perhaps, but not unheard of. Mervyn A. King, governor of the Bank of England, cited an interesting example—the Iraqi dinar—in the Ely Lecture delivered at the recent American Economics Association meeting in San Diego. Here is the story Mr. King told:

After the gulf war of 1991, Iraq was divided in two: the south ruled by Saddam Hussein, the north governed by the local Kurds. Mr. Hussein needed money to finance government spending, and in the

**Why Is That Dollar Bill in Your Pocket Worth Anything?**

*By Hal R. Varian*

Why is that dollar bill in your pocket worth anything? One answer is that it’s valuable because it says it is. To the left of the portrait of George Washington, the dollar proclaims: “This note is legal tender for all debts, public and private.”

Dollar bills are “fiat” money—they are valuable because the government in power says so. People can, however, write contracts that specify payment in other currencies. If a contract specifies payment in euros, dollars will not fulfill the contract, despite what is printed on them.

A more profound, and perhaps slightly unsettling, reason that a dollar has value is simply that lots of people are willing to accept it as payment. In this view, the value of a dollar comes not so much from government mandate as from social convention.

In the jargon of economists, the value of a dollar is a result of “network effects.” Just as a fax machine is valuable to you only if lots of other people you correspond with also have fax machines, a currency is valuable to you only if a lot of people you transact with are willing to accept it as payment.

Indeed, one can have currencies that have no government backing. Gold has been used for centuries as a medium of exchange; cigarettes were used for payment in prisoner-of-war camps in World War II; and countless other goods, including cowrie shells and peacock feathers, have functioned as money throughout history. They were money because people were willing to accept them as payment for debts, public and private. Gold, cigarettes, cowrie shells and peacock feathers all have “use value” in addition to their “exchange value.” These items were originally valued for their utility or their beauty, and they became used as currency. It is rare to see a purely paper currency functioning as money without the backing of some government or financial institution.

Rare, perhaps, but not unheard of. Mervyn A. King, governor of the Bank of England, cited an interesting example—the Iraqi dinar—in the Ely Lecture delivered at the recent American Economics Association meeting in San Diego. Here is the story Mr. King told:

After the gulf war of 1991, Iraq was divided in two: the south ruled by Saddam Hussein, the north governed by the local Kurds. Mr. Hussein needed money to finance government spending, and in the

**Monetary Lessons from Iraq**

A story from Iraq sheds light on why money has value.
time-honored tradition of dictators, created it himself.

The government could not import more of the bank notes then in use, because of United Nations sanctions, so Mr. Hussein ordered the local printing of a new currency. In May 1993, the Central Bank of Iraq announced that citizens had three weeks to exchange their old 25-dinar notes for the new “Saddam dinars,” which bore his portrait.

During the next few years, so many Saddam dinars were printed in southern Iraq that they became virtually worthless. The face value of cash in circulation rose from 22 billion dinars in 1991 to 584 billion in four years, and inflation averaged about 250 percent a year over that period.

Residents of northern Iraq could not exchange their notes. The 25-dinar notes continued to circulate and became known as the “Swiss dinars,” because they were printed with plates made in Switzerland.

The fact that the Swiss dinars continued to be used at all speaks to the power of social conventions. The Kurds in the north despised the Baghdad government, and would have much preferred to have their own currency. But there was no government in place powerful enough to mandate a currency change, so they kept using the old Swiss dinars by default.

The Swiss dinar was in fixed supply, while the Saddam dinar was flying off the printing presses, so it is not surprising that the Swiss dinar quickly became more valuable. By spring 2003, it took 300 Saddam dinars to buy one Swiss dinar.

The more interesting economic effect was the behavior of the Swiss dinar against the dollar. In fall 2002, as it became more and more likely that the United States would invade, the Swiss dinar became more and more valuable.

This appreciation was driven by expectations. If the Kurds had expected that they would once again fall under Saddam’s sway, the Swiss dinar would have quickly become worthless. As this became less likely, and the belief that future governments would accept the Swiss dinar became more widespread, the local currency became more valuable. Of course, every exchange rate movement can be interpreted in two ways: in the north, the Kurdish regional government initially interpreted the rise in the Swiss dinar against the dollar as a fall in the value of the dollar.

The government soon realized, however, that since the dollar was stable against other currencies, the correct explanation was that recounted above: the increasing belief that the Swiss dinars would, in fact, be honored by future governments.

The government was right. On July 7, 2003, the American occupation administrator, L. Paul Bremer III, announced the creation of a new Iraqi dinar that would be exchanged for the two existing currencies at a rate that implied that one Swiss dinar would be worth 150 Saddam dinars.

Interestingly, the currency markets valued the Swiss dinars somewhat higher than the official 150 exchange rate, primarily because many counterfeit 10,000-dinar Saddam notes were in circulation.

This story illustrates that paper currency can take on a life of its own, even in the absence of government backing. At the same time, it is clear that government backing makes a significant contribution to the value of paper currency: the more likely it became that the Swiss dinars would be valued by a subsequent government, the more valuable they became.


Money in the U.S. Economy

As we will see, the quantity of money circulating in the economy, called the money stock, has a powerful influence on many economic variables. But before we consider why that is true, we need to ask a preliminary question: What is the quantity of money? In particular, suppose you were given the task of measuring how much money there is in the U.S. economy. What would you include in your measure?

The most obvious asset to include is currency—the paper bills and coins in the hands of the public. Currency is clearly the most widely accepted medium of exchange in our economy. There is no doubt that it is part of the money stock.
Yet currency is not the only asset that you can use to buy goods and services. Many stores also accept personal checks. Wealth held in your checking account is almost as convenient for buying things as wealth held in your wallet. To measure the money stock, therefore, you might want to include demand deposits—balances in bank accounts that depositors can access on demand simply by writing a check.

Once you start to consider balances in checking accounts as part of the money stock, you are led to consider the large variety of other accounts that people hold at banks and other financial institutions. Bank depositors usually cannot write checks against the balances in their savings accounts, but they can easily transfer funds from savings into checking accounts. In addition, depositors in money market mutual funds can often write checks against their balances. Thus, these other accounts should plausibly be part of the U.S. money stock.

In a complex economy such as ours, it is not easy to draw a line between assets that can be called “money” and assets that cannot. The coins in your pocket clearly are part of the money stock, and the Empire State Building clearly is not, but there are many assets in between these extremes for which the choice is less clear. Because different analysts can reasonably disagree about where to draw the dividing line between monetary and nonmonetary assets, various measures of the money stock are available for the U.S. economy. Figure 1 shows the two most commonly used, designated M1 and M2. M2 includes more assets in its measure of money than does M1.

For our purposes in this book, we need not dwell on the differences between the various measures of money. None of our discussion will hinge on the distinction between M1 and M2. The important point is that the money stock for the U.S. economy includes not just currency but also deposits in banks and other financial institutions that can be readily accessed and used to buy goods and services.
WHERE IS ALL THE CURRENCY?

One puzzle about the money stock of the U.S. economy concerns the amount of currency. In 2007, there was $759 billion of currency outstanding. To put this number in perspective, we can divide it by 232 million, the number of adults (age 16 and older) in the United States. This calculation implies that the average adult holds about $3,272 of currency. Most people are surprised to learn that our economy has so much currency because they carry far less than this in their wallets.

Who is holding all this currency? No one knows for sure, but there are two plausible explanations.

The first explanation is that much of the currency is held abroad. In foreign countries without a stable monetary system, people often prefer U.S. dollars to domestic assets. It is, in fact, not unusual to see U.S. dollars used overseas as the medium of exchange, unit of account, and store of value.

The second explanation is that much of the currency is held by drug dealers, tax evaders, and other criminals. For most people in the U.S. economy, currency is not a particularly good way to hold wealth. Not only can currency be lost or stolen, but it also does not earn interest, whereas a bank deposit does. Thus, most people hold only small amounts of currency. By contrast, criminals may avoid putting their wealth in banks because a bank deposit gives police a paper trail with which to trace their illegal activities. For criminals, currency may be the best store of value available.

QUICK QUIZ List and describe the three functions of money.
Whenever an economy uses a system of fiat money, as the U.S. economy does, some agency must be responsible for regulating the system. In the United States, that agency is the Federal Reserve, often simply called the Fed. If you look at the top of a dollar bill, you will see that it is called a “Federal Reserve Note.” The Fed is an example of a central bank—an institution designed to oversee the banking system and regulate the quantity of money in the economy. Other major central banks around the world include the Bank of England, the Bank of Japan, and the European Central Bank.

**THE FED’S ORGANIZATION**

The Federal Reserve was created in 1913 after a series of bank failures in 1907 convinced Congress that the United States needed a central bank to ensure the health of the nation’s banking system. Today, the Fed is run by its board of governors, which has seven members appointed by the president and confirmed by the Senate. The governors have 14-year terms. Just as federal judges are given lifetime appointments to insulate them from politics, Fed governors are given long terms to give them independence from short-term political pressures when they formulate monetary policy.

Among the seven members of the board of governors, the most important is the chairman. The chairman directs the Fed staff, presides over board meetings, and testifies regularly about Fed policy in front of congressional committees. The president appoints the chairman to a four-year term. As this book was going to press, the chairman of the Fed was Ben Bernanke, a former economics professor who was appointed to the Fed job in 2005 by President George W. Bush.

The Federal Reserve System is made up of the Federal Reserve Board in Washington, D.C., and twelve regional Federal Reserve Banks located in major cities around the country. The presidents of the regional banks are chosen by each bank’s board of directors, whose members are typically drawn from the region’s banking and business community.

The Fed has two related jobs. The first is to regulate banks and ensure the health of the banking system. This task is largely the responsibility of the regional Federal Reserve Banks. In particular, the Fed monitors each bank’s financial condition and facilitates bank transactions by clearing checks. It also acts as a bank’s bank. That is, the Fed makes loans to banks when banks themselves want to borrow. When financially troubled banks find themselves short of cash, the Fed acts as a lender of last resort—a lender to those who cannot borrow anywhere else—to maintain stability in the overall banking system.

The Fed’s second and more important job is to control the quantity of money that is made available in the economy, called the money supply. Decisions by policymakers concerning the money supply constitute monetary policy. At the Federal Reserve, monetary policy is made by the Federal Open Market Committee (FOMC). The FOMC meets about every six weeks in Washington, D.C., to discuss the condition of the economy and consider changes in monetary policy.

**THE FEDERAL OPEN MARKET COMMITTEE**

The Federal Open Market Committee is made up of the seven members of the board of governors and five of the twelve regional bank presidents. All twelve
regional presidents attend each FOMC meeting, but only five get to vote. The five with voting rights rotate among the twelve regional presidents over time. The president of the New York Fed always gets a vote, however, because New York is the traditional financial center of the U.S. economy and because all Fed purchases and sales of government bonds are conducted at the New York Fed’s trading desk.

Through the decisions of the FOMC, the Fed has the power to increase or decrease the number of dollars in the economy. In simple metaphorical terms, you can imagine the Fed printing dollar bills and dropping them around the country by helicopter. Similarly, you can imagine the Fed using a giant vacuum cleaner to suck dollar bills out of people’s wallets. Although in practice the Fed’s methods for changing the money supply are more complex and subtle than this, the helicopter-vacuum metaphor is a good first step to understanding the meaning of monetary policy.

Later in this chapter, we discuss how the Fed actually changes the money supply, but it is worth noting here that the Fed’s primary tool is the open-market operation—the purchase and sale of U.S. government bonds. (Recall that a U.S. government bond is a certificate of indebtedness of the federal government.) If the FOMC decides to increase the money supply, the Fed creates dollars and uses them to buy government bonds from the public in the nation’s bond markets. After the purchase, these dollars are in the hands of the public. Thus, an open-market purchase of bonds by the Fed increases the money supply. Conversely, if the FOMC decides to decrease the money supply, the Fed sells government bonds from its portfolio to the public in the nation’s bond markets. After the sale, the dollars it receives for the bonds are out of the hands of the public. Thus, an open-market sale of bonds by the Fed decreases the money supply.

Central banks are important institutions because changes in the money supply can profoundly affect the economy. One of the Ten Principles of Economics in Chapter 1 is that prices rise when the government prints too much money. Another of the Ten Principles of Economics is that society faces a short-run trade-off between inflation and unemployment. The power of the Fed rests on these principles. For reasons we discuss more fully in the coming chapters, the Fed’s policy decisions have an important influence on the economy’s rate of inflation in the long run and the economy’s employment and production in the short run. Indeed, the chairman of the Federal Reserve has been called the second most powerful person in the United States.

**Quick Quiz** What are the primary responsibilities of the Federal Reserve? If the Fed wants to increase the supply of money, how does it usually do so?

---

**Banks and the Money Supply**

So far, we have introduced the concept of “money” and discussed how the Federal Reserve controls the supply of money by buying and selling government bonds in open-market operations. Although this explanation of the money supply is correct, it is not complete. In particular, it omits the central role that banks play in the monetary system.

Recall that the amount of money you hold includes both currency (the bills in your wallet and coins in your pocket) and demand deposits (the balance in your checking account). Because demand deposits are held in banks, the behavior of
banks can influence the quantity of demand deposits in the economy and, therefore, the money supply. This section examines how banks affect the money supply and, in doing so, how they complicate the Fed’s job of controlling the money supply.

The Simple Case of 100-Percent-Reserve Banking

To see how banks influence the money supply, let’s first imagine a world without any banks at all. In this simple world, currency is the only form of money. To be concrete, let’s suppose that the total quantity of currency is $100. The supply of money is, therefore, $100.

Now suppose that someone opens a bank, appropriately called First National Bank. First National Bank is only a depository institution—that is, it accepts deposits but does not make loans. The purpose of the bank is to give depositors a safe place to keep their money. Whenever a person deposits some money, the bank keeps the money in its vault until the depositor comes to withdraw it or writes a check against his or her balance. Deposits that banks have received but have not loaned out are called reserves. In this imaginary economy, all deposits are held as reserves, so this system is called 100-percent-reserve banking.

We can express the financial position of First National Bank with a T-account, which is a simplified accounting statement that shows changes in a bank’s assets and liabilities. Here is the T-account for First National Bank if the economy’s entire $100 of money is deposited in the bank:

<table>
<thead>
<tr>
<th>FIRST NATIONAL BANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assets</td>
</tr>
<tr>
<td>Reserves</td>
</tr>
</tbody>
</table>

On the left side of the T-account are the bank’s assets of $100 (the reserves it holds in its vaults). On the right side are the bank’s liabilities of $100 (the amount it owes to its depositors). Notice that the assets and liabilities of First National Bank exactly balance.

Now consider the money supply in this imaginary economy. Before First National Bank opens, the money supply is the $100 of currency that people are holding. After the bank opens and people deposit their currency, the money supply is the $100 of demand deposits. (There is no longer any currency outstanding, for it is all in the bank vault.) Each deposit in the bank reduces currency and raises demand deposits by exactly the same amount, leaving the money supply unchanged. Thus, if banks hold all deposits in reserve, banks do not influence the supply of money.

Money Creation with Fractional-Reserve Banking

Eventually, the bankers at First National Bank may start to reconsider their policy of 100-percent-reserve banking. Leaving all that money idle in their vaults seems unnecessary. Why not lend some of it out and earn a profit by charging interest on the loans? Families buying houses, firms building new factories, and students paying for college would all be happy to pay interest to borrow some of that money for a while. First National Bank has to keep some reserves so that cur-
currency is available if depositors want to make withdrawals. But if the flow of new deposits is roughly the same as the flow of withdrawals, First National needs to keep only a fraction of its deposits in reserve. Thus, First National adopts a system called **fractional-reserve banking**.

The fraction of total deposits that a bank holds as reserves is called the **reserve ratio**. This ratio is determined by a combination of government regulation and bank policy. As we discuss more fully later in the chapter, the Fed sets a minimum amount of reserves that banks must hold, called a **reserve requirement**. In addition, banks may hold reserves above the legal minimum, called **excess reserves**, so they can be more confident that they will not run short of cash. For our purpose here, we take the reserve ratio as given to examine what fractional-reserve banking means for the money supply.

Let’s suppose that First National has a reserve ratio of 1/10, or 10 percent. This means that it keeps 10 percent of its deposits in reserve and loans out the rest. Now let’s look again at the bank’s T-account:

<table>
<thead>
<tr>
<th>FIRST NATIONAL BANK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assets</strong></td>
</tr>
<tr>
<td>Reserves $10.00</td>
</tr>
<tr>
<td>Loans 90.00</td>
</tr>
</tbody>
</table>

First National still has $100 in liabilities because making the loans did not alter the bank’s obligation to its depositors. But now the bank has two kinds of assets: It has $10 of reserves in its vault, and it has loans of $90. (These loans are liabilities of the people taking out the loans, but they are assets of the bank making the loans because the borrowers will later repay the bank.) In total, First National’s assets still equal its liabilities.

Once again consider the supply of money in the economy. Before First National makes any loans, the money supply is the $100 of deposits in the bank. Yet when First National makes these loans, the money supply increases. The depositors still have demand deposits totaling $100, but now the borrowers hold $90 in currency. The money supply (which equals currency plus demand deposits) equals $190. Thus, **when banks hold only a fraction of deposits in reserve, banks create money**.

At first, this creation of money by fractional-reserve banking may seem too good to be true because it appears that the bank has created money out of thin air. To make this creation of money seem less miraculous, note that when First National Bank loans out some of its reserves and creates money, it does not create any wealth. Loans from First National give the borrowers some currency and thus the ability to buy goods and services. Yet the borrowers are also taking on debts, so the loans do not make them any richer. In other words, as a bank creates the asset of money, it also creates a corresponding liability for its borrowers. At the end of this process of money creation, the economy is more liquid in the sense that there is more of the medium of exchange, but the economy is no wealthier than before.

**The Money Multiplier**

The creation of money does not stop with First National Bank. Suppose the borrower from First National uses the $90 to buy something from someone who then...
deposits the currency in Second National Bank. Here is the T-account for Second National Bank:

**SECOND NATIONAL BANK**

<table>
<thead>
<tr>
<th>Assets</th>
<th>Liabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserves</td>
<td>$ 9.00</td>
</tr>
<tr>
<td>Loans</td>
<td>$ 81.00</td>
</tr>
</tbody>
</table>

After the deposit, this bank has liabilities of $90. If Second National also has a reserve ratio of 10 percent, it keeps assets of $9 in reserve and makes $81 in loans. In this way, Second National Bank creates an additional $81 of money. If this $81 is eventually deposited in Third National Bank, which also has a reserve ratio of 10 percent, this bank keeps $8.10 in reserve and makes $72.90 in loans. Here is the T-account for Third National Bank:

**THIRD NATIONAL BANK**

<table>
<thead>
<tr>
<th>Assets</th>
<th>Liabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserves</td>
<td>$ 8.10</td>
</tr>
<tr>
<td>Loans</td>
<td>$ 72.90</td>
</tr>
</tbody>
</table>

The process goes on and on. Each time that money is deposited and a bank loan is made, more money is created.

How much money is eventually created in this economy? Let’s add it up:

Original deposit = $100.00
First National lending = $ 90.00 [ = .9 × $100.00]
Second National lending = $ 81.00 [ = .9 × $90.00]
Third National lending = $ 72.90 [ = .9 × $81.00]

• • •

Total money supply = $1,000.00

It turns out that even though this process of money creation can continue forever, it does not create an infinite amount of money. If you laboriously add the infinite sequence of numbers in the foregoing example, you find the $100 of reserves generates $1,000 of money. The amount of money the banking system generates with each dollar of reserves is called the **money multiplier**. In this imaginary economy, where the $100 of reserves generates $1,000 of money, the money multiplier is 10.

What determines the size of the money multiplier? It turns out that the answer is simple: The **money multiplier is the reciprocal of the reserve ratio**. If $R$ is the reserve ratio for all banks in the economy, then each dollar of reserves generates $1/R$ dollars of money. In our example, $R = 1/10$, so the money multiplier is 10.

This reciprocal formula for the money multiplier makes sense. If a bank holds $1,000 in deposits, then a reserve ratio of $1/10$ (10 percent) means that the bank must hold $100 in reserves. The money multiplier just turns this idea around: If the banking system as a whole holds a total of $100 in reserves, it can have only $1,000 in deposits. In other words, if $R$ is the ratio of reserves to deposits at each
bank (that is, the reserve ratio), then the ratio of deposits to reserves in the banking system (that is, the money multiplier) must be $1/R$.

This formula shows how the amount of money banks create depends on the reserve ratio. If the reserve ratio were only 1/20 (5 percent), then the banking system would have 20 times as much in deposits as in reserves, implying a money multiplier of 20. Each dollar of reserves would generate $20 of money. Similarly, if the reserve ratio were 1/5 (20 percent), deposits would be 5 times reserves, the money multiplier would be 5, and each dollar of reserves would generate $5 of money. Thus, the higher the reserve ratio, the less of each deposit banks loan out, and the smaller the money multiplier. In the special case of 100-percent-reserve banking, the reserve ratio is 1, the money multiplier is 1, and banks do not make loans or create money.

**THE FED’S TOOLS OF MONETARY CONTROL**

As we have already discussed, the Federal Reserve is responsible for controlling the supply of money in the economy. Now that we understand how fractional-reserve banking works, we are in a better position to understand how the Fed carries out this job. Because banks create money in a system of fractional-reserve banking, the Fed’s control of the money supply is indirect. When the Fed decides to change the money supply, it must consider how its actions will work through the banking system.

The Fed has three tools in its monetary toolbox: open-market operations, reserve requirements, and the discount rate. Let’s discuss how the Fed uses each of these tools.

**Open-Market Operations**  As we noted earlier, the Fed conducts open-market operations when it buys or sells government bonds. To increase the money supply, the Fed instructs its bond traders at the New York Fed to buy bonds from the public in the nation’s bond markets. The dollars the Fed pays for the bonds increase the number of dollars in the economy. Some of these new dollars are held as currency, and some are deposited in banks. Each new dollar held as currency increases the money supply by exactly $1. Each new dollar deposited in a bank increases the money supply by more than a dollar because it increases reserves and, thereby, the amount of money that the banking system can create.

To reduce the money supply, the Fed does just the opposite: It sells government bonds to the public in the nation’s bond markets. The public pays for these bonds with its holdings of currency and bank deposits, directly reducing the amount of money in circulation. In addition, as people make withdrawals from banks to buy these bonds from the Fed, banks find themselves with a smaller quantity of reserves. In response, banks reduce the amount of lending, and the process of money creation reverses itself.

Open-market operations are easy to conduct. In fact, the Fed’s purchases and sales of government bonds in the nation’s bond markets are similar to the transactions that any individual might undertake for his own portfolio. (Of course, when an individual buys or sells a bond, money changes hands, but the amount of money in circulation remains the same.) In addition, the Fed can use open-market operations to change the money supply by a small or large amount on any day without major changes in laws or bank regulations. Therefore, open-market operations are the tool of monetary policy that the Fed uses most often.
Reserve Requirements

The Fed also influences the money supply with reserve requirements, which are regulations on the minimum amount of reserves that banks must hold against deposits. Reserve requirements influence how much money the banking system can create with each dollar of reserves. An increase in reserve requirements means that banks must hold more reserves and, therefore, can loan out less of each dollar that is deposited; as a result, it raises the reserve ratio, lowers the money multiplier, and decreases the money supply. Conversely, a decrease in reserve requirements lowers the reserve ratio, raises the money multiplier, and increases the money supply.

The Fed uses changes in reserve requirements only rarely because these changes disrupt the business of banking. When the Fed increases reserve requirements, for instance, some banks find themselves short of reserves, even though they have

Bernanke’s Creativity

As one of the nation’s most distinguished monetary economists, Federal Reserve Chairman Ben Bernanke spent a quarter century studying the Fed’s role in the Great Depression before joining the Federal Reserve Board as a governor in 2002. At a November 2002 conference celebrating Milton Friedman’s 90th birthday, here is how Mr. Bernanke concluded his scholarly speech about the Great Depression and Mr. Friedman’s analysis of that epochal event: “Let me end my talk by abusing slightly my status as an official representative of the Federal Reserve. I would like to say to Milton and Anna [Schwartz, the coauthors of the classic A Monetary History of the United States, 1863–1960]: Regarding the Great Depression. You’re right, we [at the Fed] did it. We’re very sorry. But thanks to you, we won’t do it again.”
To appreciate the magnitude of the Fed’s actions, let’s review how Mr. Bernanke and his colleagues have ratcheted up their innovative policy responses as the credit crisis has deepened. Keep in mind that these often-revolutionary actions are above and beyond the Fed’s extraordinary response via its traditional monetary-policy channel, through which the Fed has reduced the inflation-adjusted overnight fed-funds rate by a staggering 5 percentage points (from 3.25 percent in September to a negative 1.75 percent today) over a six-month period.

The financial and credit markets first encountered severe strain in the middle of last summer. However, on the heels of economic reports showing that the economy had expanded at a 3.4 percent annual rate during last year’s second quarter and more than 400,000 jobs had been created during the previous three months, the Fed decided at its Aug. 7 meeting not to lower the fed-funds rate, which is the interest rate commercial banks charge each other for mostly overnight loans. Markets revolted. The Fed tried to soothe their nerves on Aug. 10 by issuing a press release inviting “depository institutions” (i.e., commercial banks) to borrow funds directly from the Fed at its discount window.

Commercial banks avoided the discount window because of the stigma attached to it. So, the Fed on Dec. 12 announced an extraordinary policy (a “temporary Term Auction Facility” (TAF)) whereby banks could bid twice a month for 28- to 35-day loans by pledging the kinds of collateral acceptable at the discount window.

On March 11, the Fed established yet another unorthodox mechanism (the Term Securities Lending Facility (TSLF)) to inject liquidity into the overstressed financial markets. In an unprecedented move, the Fed would begin lending up to $200 billion in Treasury securities for 28 days to primary dealers (i.e., investment banks), which could use so-called Schedule 1 securities as collateral.

On March 14, in yet another action that had not been conducted since the Great Depression, the Fed agreed to provide a 28-day loan to JPMorgan Chase, which would then loan the funds to the cash-strapped Bear Stearns investment bank.

Two days later, on Sunday, March 16, to prevent the imminent bankruptcy of Bear Stearns, the Fed approved JPMorgan’s takeover of Bear for $2 a share, which was later raised to $10. (Last year Bear Stearns sold for more than $150 a share.) To induce the takeover, the Fed has agreed to absorb losses up to $29 billion on questionable securities held by Bear Stearns, putting taxpayer funds at risk. In yet another breathtaking move, the Fed also opened its discount-window lending operation to primary-dealer investment banks.

These are not normal times, and the Fed clearly has been reacting accordingly. Whether its unprecedented actions will be enough to ameliorate the credit crisis and prevent a deep recession remains to be seen. Mr. Bernanke’s promise that “we won’t do it again” will not be broken for lack of effort or creativity.


seen no change in deposits. As a result, they have to curtail lending until they build their level of reserves to the new required level.

**The Discount Rate** The third tool in the Fed’s toolbox is the **discount rate**, the interest rate on the loans that the Fed makes to banks. A bank borrows from the Fed when it has too few reserves to meet its reserve requirements. This might occur because the bank has made too many loans or because it has experienced unexpectedly high withdrawals. When the Fed makes such a loan to a bank, the banking system has more reserves than it otherwise would, and these additional reserves allow the banking system to create more money.

The Fed can alter the money supply by changing the discount rate. A higher discount rate discourages banks from borrowing reserves from the Fed. Thus, an **discount rate** the interest rate on the loans that the Fed makes to banks
increase in the discount rate reduces the quantity of reserves in the banking system, which in turn reduces the money supply. Conversely, a lower discount rate encourages banks to borrow from the Fed, increasing the quantity of reserves and the money supply.

The Fed uses discount lending not only to control the money supply but also to help financial institutions when they are in trouble. For example, when the stock market crashed by 22 percent on October 19, 1987, many Wall Street brokerage firms found themselves temporarily in need of funds to finance the high volume of stock trading. The next morning, before the stock market opened, Fed Chairman Alan Greenspan announced the Fed’s “readiness to serve as a source of liquidity to support the economic and financial system.” Many economists believe that Greenspan’s reaction to the stock crash was an important reason it had so few repercussions.

Similarly, in 2007 and 2008, a fall in house prices throughout the United States led to a sharp rise in the number of homeowners defaulting on their mortgage loans, and many financial institutions holding those mortgages ran into trouble. In an attempt to prevent these events from having broader economic ramifications, the Fed provided loans to many of the financial institutions in distress. The details are described in the In The News box on the previous two pages.

**Problems in Controlling the Money Supply**

The Fed’s three tools—open-market operations, reserve requirements, and the discount rate—have powerful effects on the money supply. Yet the Fed’s control of the money supply is not precise. The Fed must wrestle with two problems, each of which arises because much of the money supply is created by our system of fractional-reserve banking.

The first problem is that the Fed does not control the amount of money that households choose to hold as deposits in banks. The more money households deposit, the more reserves banks have, and the more money the banking system can create. And the less money households deposit, the less reserves banks have, and the less money the banking system can create. To see why this is a problem, suppose that one day people begin to lose confidence in the banking system and, therefore, decide to withdraw deposits and hold more currency. When this happens, the banking system loses reserves and creates less money. The money supply falls, even without any Fed action.

The second problem of monetary control is that the Fed does not control the amount that bankers choose to lend. When money is deposited in a bank, it creates more money only when the bank loans it out. Because banks can choose to hold excess reserves instead, the Fed cannot be sure how much money the banking system will create. For instance, suppose that one day bankers become more cautious about economic conditions and decide to make fewer loans and hold greater reserves. In this case, the banking system creates less money than it otherwise would. Because of the bankers’ decision, the money supply falls.

Hence, in a system of fractional-reserve banking, the amount of money in the economy depends in part on the behavior of depositors and bankers. Because the Fed cannot control or perfectly predict this behavior, it cannot perfectly control the money supply. Yet if the Fed is vigilant, these problems need not be large. The Fed collects data on deposits and reserves from banks every week, so it is quickly aware of any changes in depositor or banker behavior. It can, therefore, respond to these changes and keep the money supply close to whatever level it chooses.
BANK RUNS AND THE MONEY SUPPLY

Although you have probably never witnessed a bank run in real life, you may have seen one depicted in movies such as *Mary Poppins* or *It’s a Wonderful Life*. A bank run occurs when depositors suspect that a bank may go bankrupt and, therefore, “run” to the bank to withdraw their deposits. The United States has not seen a major bank run in recent history, but in the United Kingdom, a bank called Northern Rock experienced a run in 2007 and, as a result, was eventually taken over by the government.

Bank runs are a problem for banks under fractional-reserve banking. Because a bank holds only a fraction of its deposits in reserve, it cannot satisfy withdrawal requests from all depositors. Even if the bank is in fact solvent (meaning that its assets exceed its liabilities), it will not have enough cash on hand to allow all depositors immediate access to all of their money. When a run occurs, the bank is forced to close its doors until some bank loans are repaid or until some lender of last resort (such as the Fed) provides it with the currency it needs to satisfy depositors.

Bank runs complicate the control of the money supply. An important example of this problem occurred during the Great Depression in the early 1930s. After a wave of bank runs and bank closings, households and bankers became more cautious. Households withdrew their deposits from banks, preferring to hold their money in the form of currency. This decision reversed the process of money creation, as bankers responded to falling reserves by reducing bank loans. At the same time, bankers increased their reserve ratios so that they would have enough cash on hand to meet their depositors’ demands in any future bank runs. The higher reserve ratio reduced the money multiplier, which further reduced the money supply. From 1929 to 1933, the money supply fell by 28 percent, even without the Federal Reserve taking any deliberate contractionary action. Many economists point to this massive fall in the money supply to explain the high unemployment and falling prices that prevailed during this period. (In future chapters, we examine the mechanisms by which changes in the money supply affect unemployment and prices.)

Today, bank runs are not a major problem for the U.S. banking system or the Fed. The federal government now guarantees the safety of deposits at most banks, primarily through the Federal Deposit Insurance Corporation (FDIC). Depositors do not run on their banks because they are confident that, even if their bank goes bankrupt, the FDIC will make good on the deposits. The policy of government deposit insurance has costs: Bankers whose deposits are guaranteed may have too little incentive to avoid bad risks when making loans. But one benefit of deposit insurance is a more stable banking system. As a result, most people see bank runs only in the movies.

THE FEDERAL FUNDS RATE

If you read about U.S. monetary policy in the newspaper, you will find much discussion of the federal funds rate. This raises several questions:

Q: What is the federal funds rate?
A: The **federal funds rate** is the short-term interest rate that banks charge one another for loans. If one bank finds itself short of reserves while another
bank has excess reserves, the second bank can lend some reserves to the first. The loans are temporary—typically overnight. The price of the loan is the federal funds rate.

Q: Does the federal funds rate matter only for banks?
A: Not at all. While only banks borrow directly in the federal funds market, the economic impact of this market is much broader. Because different parts of the financial system are highly interconnected, interest rates on different kinds of loans are strongly correlated with one another. So when the federal funds rate rises or falls, other interest rates often move in the same direction.

Q: What does the Federal Reserve have to do with the federal funds rate?
A: In recent years, the Federal Reserve has set a target goal for the federal funds rate. When the Federal Open Market Committee meets approximately every six weeks, it decides whether to raise or lower that target.

Q: How can the Fed make the federal funds rate hit the target it sets?
A: Although the actual federal funds rate is set by supply and demand in the market for loans among banks, the Fed can use open-market operations to influence that market. For example, when the Fed buys bonds in open-market operations, it injects reserves into the banking system. With more reserves in the system, fewer banks find themselves in need of borrowing reserves to meet reserve requirements. The fall in demand for borrowing reserves decreases the price of such borrowing, which is the federal funds rate. Conversely, when the Fed sells bonds and withdraws reserves from the banking system, more banks find themselves short of reserves, and they bid up the price of borrowing reserves. Thus, open-market purchases lower the federal funds rate, and open-market sales raise the federal funds rate.

Q: But don’t these open-market operations affect the money supply?
A: Yes, absolutely. When the Fed announces a change in the federal funds rate, it is committing itself to the open-market operations necessary to make that change happen, and these open-market operations will alter the supply of money. Decisions by the FOMC to change the target for the federal funds rate are also decisions to change the money supply. They are two sides of the same coin. Other things equal, a decrease in the target for the federal funds rate means an expansion in the money supply, and an increase in the target for the federal funds rate means a contraction in the money supply.

**Quick Quiz** Describe how banks create money. • If the Fed wanted to use all three of its policy tools to decrease the money supply, what would it do?

**CONCLUSION**

Some years ago, a book made the best-seller list with the title *Secrets of the Temple: How the Federal Reserve Runs the Country*. Although no doubt an exaggeration, this title did highlight the important role of the monetary system in our daily lives. Whenever we buy or sell anything, we are relying on the extraordinarily useful social convention called “money.” Now that we know what money is and what determines its supply, we can discuss how changes in the quantity of money affect the economy. We begin to address that topic in the next chapter.
SUMMARY

• The term money refers to assets that people regularly use to buy goods and services.

• Money serves three functions. As a medium of exchange, it provides the item used to make transactions. As a unit of account, it provides the way in which prices and other economic values are recorded. As a store of value, it provides a way of transferring purchasing power from the present to the future.

• Commodity money, such as gold, is money that has intrinsic value: It would be valued even if it were not used as money. Fiat money, such as paper dollars, is money without intrinsic value: It would be worthless if it were not used as money.

• In the U.S. economy, money takes the form of currency and various types of bank deposits, such as checking accounts.

• The Federal Reserve, the central bank of the United States, is responsible for regulating the U.S. monetary system. The Fed chairman is appointed by the president and confirmed by Congress every 4 years. The chairman is the lead member of the Federal Open Market Committee, which meets about every 6 weeks to consider changes in monetary policy.

• The Fed controls the money supply primarily through open-market operations: The purchase of government bonds increases the money supply, and the sale of government bonds decreases the money supply. The Fed can also expand the money supply by lowering reserve requirements or decreasing the discount rate, and it can contract the money supply by raising reserve requirements or increasing the discount rate.

• When banks loan out some of their deposits, they increase the quantity of money in the economy. Because banks influence the money supply in this way, the Fed’s control of the money supply is imperfect.

• The Federal Reserve has in recent years set monetary policy by choosing a target for the federal funds rate, a short-term interest rate at which banks make loans to one another. As the Fed achieves its target, it adjusts the money supply.

KEY CONCEPTS

money, p. 458
medium of exchange, p. 459
unit of account, p. 459
store of value, p. 459
liquidity, p. 459
commodity money, p. 459
fiat money, p. 460
currency, p. 461
demand deposits, p. 462
Federal Reserve (Fed), p. 464
central bank, p. 464
money supply, p. 464
monetary policy, p. 464
reserves, p. 466
fractional-reserve banking, p. 467
reserve ratio, p. 467
money multiplier, p. 468
open-market operations, p. 469
reserve requirements, p. 470
discount rate, p. 471
federal funds rate, p. 473
QUESTIONS FOR REVIEW

1. What distinguishes money from other assets in the economy?
2. What is commodity money? What is fiat money? Which kind do we use?
3. What are demand deposits and why should they be included in the stock of money?
4. Who is responsible for setting monetary policy in the United States? How is this group chosen?
5. If the Fed wants to increase the money supply with open-market operations, what does it do?
6. Why don’t banks hold 100 percent reserves? How is the amount of reserves banks hold related to the amount of money the banking system creates?
7. What is the discount rate? What happens to the money supply when the Fed raises the discount rate?
8. What are reserve requirements? What happens to the money supply when the Fed raises reserve requirements?
9. Why can’t the Fed control the money supply perfectly?

PROBLEMS AND APPLICATIONS

1. Which of the following are money in the U.S. economy? Which are not? Explain your answers by discussing each of the three functions of money.
   a. a U.S. penny
   b. a Mexican peso
   c. a Picasso painting
   d. a plastic credit card
2. What characteristics of an asset make it useful as a medium of exchange? As a store of value?
3. Go to the website of the Federal Reserve Bank of St. Louis (http://www.stlouisfed.org) to find some information about the Fed. Find a map of the Federal Reserve districts. If you live in the United States, find what district you live in. Where is the Federal Reserve Bank for your district located? (Extra credit: What state has two Federal Reserve Banks?)
4. Your uncle repays a $100 loan from Tenth National Bank (TNB) by writing a $100 check from his TNB checking account. Use T-accounts to show the effect of this transaction on your uncle and on TNB. Has your uncle’s wealth changed? Explain.
5. Beleaguered State Bank (BSB) holds $250 million in deposits and maintains a reserve ratio of 10 percent.
   a. Show a T-account for BSB.
   b. Now suppose that BSB’s largest depositor withholds $10 million in cash from her account. If BSB decides to restore its reserve ratio by reducing the amount of loans outstanding, show its new T-account.
   c. Explain what effect BSB’s action will have on other banks.
   d. Why might it be difficult for BSB to take the action described in part (b)? Discuss another way for BSB to return to its original reserve ratio.
6. You take $100 you had kept under your mattress and deposit it in your bank account. If this $100 stays in the banking system as reserves and if banks hold reserves equal to 10 percent of deposits, by how much does the total amount of deposits in the banking system increase? By how much does the money supply increase?
7. The Federal Reserve conducts a $10 million open-market purchase of government bonds. If the required reserve ratio is 10 percent, what is the largest possible increase in the money supply that could result? Explain. What is the smallest possible increase? Explain.
8. Assume that the reserve requirement is 5 percent. All other things equal, will the money supply expand more if the Federal Reserve buys $2,000 worth of bonds or if someone deposits in a bank $2,000 that he had been hiding in his cookie jar? If one creates more, how much more does it create? Support your thinking.
9. Suppose that the T-account for First National Bank is as follows:
in reserves and the change in the money supply?

12. Assume that the reserve requirement is 20%.
Also assume that banks do not hold excess reserves and there is no cash held by the public. The Federal Reserve decides that it wants to expand the money supply by $40 million dollars.

a. If the Fed is using open-market operations, will it buy or sell bonds?
b. What quantity of bonds does the Fed need to buy or sell to accomplish the goal? Explain your reasoning.

13. The economy of El mendyn contains 2,000 $1 bills.

a. If people hold all money as currency, what is the quantity of money?
b. If people hold all money as demand deposits and banks maintain 100 percent reserves, what is the quantity of money?
c. If people hold equal amounts of currency and demand deposits and banks maintain 100 percent reserves, what is the quantity of money?
d. If people hold all money as demand deposits and banks maintain a reserve ratio of 10 percent, what is the quantity of money?
e. If people hold equal amounts of currency and demand deposits and banks maintain a reserve ratio of 10 percent, what is the quantity of money?

<table>
<thead>
<tr>
<th>Assets</th>
<th>Liabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserves</td>
<td>Deposits $500,000</td>
</tr>
<tr>
<td>Loans</td>
<td>$100,000</td>
</tr>
<tr>
<td>Loans</td>
<td>$400,000</td>
</tr>
</tbody>
</table>

a. If the Fed requires banks to hold 5 percent of deposits as reserves, how much in excess reserves does First National now hold?
b. Assume that all other banks hold only the required amount of reserves. If First National decides to reduce its reserves to only the required amount, by how much would the economy’s money supply increase?

10. Suppose that the reserve requirement for checking deposits is 10 percent and that banks do not hold any excess reserves.

a. If the Fed sells $1 million of government bonds, what is the effect on the economy’s reserves and money supply?
b. Now suppose the Fed lowers the reserve requirement to 5 percent, but banks choose to hold another 5 percent of deposits as excess reserves. Why might banks do so? What is the overall change in the money multiplier and the money supply as a result of these actions?

11. Assume that the banking system has total reserves of $100 billion. Assume also that required reserves are 10 percent of checking deposits and that banks hold no excess reserves and households hold no currency.

a. What is the money multiplier? What is the money supply?
b. If the Fed now raises required reserves to 20 percent of deposits, what is the change in reserves and the change in the money supply?
Money Growth and Inflation

Todays, if you want to buy an ice-cream cone, you'll need at least a couple of dollars, but that has not always been the case. In the 1930s, my grandmother ran a sweet shop in Trenton, New Jersey, and she sold ice-cream cones in two sizes. A cone with a small scoop of ice cream cost three cents. Hungry customers could buy a large scoop for a nickel.

You may not be surprised at the increase in the price of ice cream. In our economy, most prices tend to rise over time. This increase in the overall level of prices is called inflation. Earlier in the book, we examined how economists measure the inflation rate as the percentage change in the consumer price index (CPI), the GDP deflator, or some other index of the overall price level. These price indexes show that, over the past 70 years, prices have risen on average about 4 percent per year. Accumulated over so many years, a 4 percent annual inflation rate leads to a sixteenfold increase in the price level.

Inflation may seem natural and inevitable to a person who grew up in the United States during recent decades, but in fact, it is not inevitable at all. There were long periods in the 19th century during which most prices fell—a phenomenon called deflation. The average level of prices in the U.S. economy was 23 percent lower in 1896 than in 1880, and this deflation was a major issue in the presidential election of 1896. Farmers, who had accumulated large debts, suffered when the fall in crop prices reduced their incomes and thus their ability to pay off their debts. They advocated government policies to reverse the deflation.
Although inflation has been the norm in more recent history, there has been substantial variation in the rate at which prices rise. During the 1990s, prices rose at an average rate of about 2 percent per year. By contrast, in the 1970s, prices rose by 7 percent per year, which meant a doubling of the price level over the decade. The public often views such high rates of inflation as a major economic problem. In fact, when President Jimmy Carter ran for reelection in 1980, challenger Ronald Reagan pointed to high inflation as one of the failures of Carter’s economic policy.

International data show an even broader range of inflation experiences. In 2007, while the U.S. inflation rate was about 4 percent, inflation was 0.7 percent in Japan, 13 percent in Russia, and 25 percent in Venezuela. And even the high inflation rates in Russia and Venezuela are moderate by some standards. In February 2008, the central bank of Zimbabwe announced the inflation rate in its economy had reached 24,000 percent, while some independent estimates put the figure even higher. An extraordinarily high rate of inflation such as this is called hyperinflation.

What determines whether an economy experiences inflation and, if so, how much? This chapter answers this question by developing the quantity theory of money. Chapter 1 summarized this theory as one of the Ten Principles of Economics: Prices rise when the government prints too much money. This insight has a long and venerable tradition among economists. The quantity theory was discussed by the famous 18th-century philosopher and economist David Hume and has been advocated more recently by the prominent economist Milton Friedman. This theory can explain moderate inflations, such as those we have experienced in the United States, as well as hyperinflations.

After developing a theory of inflation, we turn to a related question: Why is inflation a problem? At first glance, the answer to this question may seem obvious: Inflation is a problem because people don’t like it. In the 1970s, when the United States experienced a relatively high rate of inflation, opinion polls placed inflation as the most important issue facing the nation. President Ford echoed this sentiment in 1974 when he called inflation “public enemy number one.” Ford wore a “WIN” button on his lapel—for Whip Inflation Now.

But what, exactly, are the costs that inflation imposes on a society? The answer may surprise you. Identifying the various costs of inflation is not as straightforward as it first appears. As a result, although all economists decry hyperinflation, some economists argue that the costs of moderate inflation are not nearly as large as the public believes.

**THE CLASSICAL THEORY OF INFLATION**

We begin our study of inflation by developing the quantity theory of money. This theory is often called “classical” because it was developed by some of the earliest economic thinkers. Most economists today rely on this theory to explain the long-run determinants of the price level and the inflation rate.

**The Level of Prices and the Value of Money**

Suppose we observe over some period of time the price of an ice-cream cone rising from a nickel to a dollar. What conclusion should we draw from the fact that
people are willing to give up so much more money in exchange for a cone? It is possible that people have come to enjoy ice cream more (perhaps because some chemist has developed a miraculous new flavor). Yet that is probably not the case. It is more likely that people’s enjoyment of ice cream has stayed roughly the same and that, over time, the money used to buy ice cream has become less valuable. Indeed, the first insight about inflation is that it is more about the value of money than about the value of goods.

This insight helps point the way toward a theory of inflation. When the consumer price index and other measures of the price level rise, commentators are often tempted to look at the many individual prices that make up these price indexes: “The CPI rose by 3 percent last month, led by a 20 percent rise in the price of coffee and a 30 percent rise in the price of heating oil.” Although this approach does contain some interesting information about what’s happening in the economy, it also misses a key point: Inflation is an economy-wide phenomenon that concerns, first and foremost, the value of the economy’s medium of exchange.

The economy’s overall price level can be viewed in two ways. So far, we have viewed the price level as the price of a basket of goods and services. When the price level rises, people have to pay more for the goods and services they buy. Alternatively, we can view the price level as a measure of the value of money. A rise in the price level means a lower value of money because each dollar in your wallet now buys a smaller quantity of goods and services.

It may help to express these ideas mathematically. Suppose $P$ is the price level as measured, for instance, by the consumer price index or the GDP deflator. Then $P$ measures the number of dollars needed to buy a basket of goods and services. Now turn this idea around: The quantity of goods and services that can be bought with $1 equals $1/P$. In other words, if $P$ is the price of goods and services measured in terms of money, $1/P$ is the value of money measured in terms of goods and services. Thus, when the overall price level rises, the value of money falls.

**Money Supply, Money Demand, and Monetary Equilibrium**

What determines the value of money? The answer to this question, like many in economics, is supply and demand. Just as the supply and demand for bananas determines the price of bananas, the supply and demand for money determines the value of money. Thus, our next step in developing the quantity theory of money is to consider the determinants of money supply and money demand.

First consider money supply. In the preceding chapter, we discussed how the Federal Reserve, together with the banking system, determines the supply of money. When the Fed sells bonds in open-market operations, it receives dollars in exchange and contracts the money supply. When the Fed buys government bonds, it pays out dollars and expands the money supply. In addition, if any of these dollars are deposited in banks which hold some as reserves and loan out the rest, the money multiplier swings into action, and these open-market operations can have an even greater effect on the money supply. For our purposes in this chapter, we ignore the complications introduced by the banking system and simply take the quantity of money supplied as a policy variable that the Fed controls.

Now consider money demand. Most fundamentally, the demand for money reflects how much wealth people want to hold in liquid form. Many factors influence the quantity of money demanded. The amount of currency that people hold...
in their wallets, for instance, depends on how much they rely on credit cards and on whether an automatic teller machine is easy to find. And as we will emphasize in Chapter 24, the quantity of money demanded depends on the interest rate that a person could earn by using the money to buy an interest-bearing bond rather than leaving it in a wallet or low-interest checking account.

Although many variables affect the demand for money, one variable stands out in importance: the average level of prices in the economy. People hold money because it is the medium of exchange. Unlike other assets, such as bonds or stocks, people can use money to buy the goods and services on their shopping lists. How much money they choose to hold for this purpose depends on the prices of those goods and services. The higher prices are, the more money the typical transaction requires, and the more money people will choose to hold in their wallets and checking accounts. That is, a higher price level (a lower value of money) increases the quantity of money demanded.

What ensures that the quantity of money the Fed supplies balances the quantity of money people demand? The answer, it turns out, depends on the time horizon being considered. Later in this book, we examine the short-run answer and learn that interest rates play a key role. In the long run, however, the answer is different and much simpler. In the long run, the overall level of prices adjusts to the level at which the demand for money equals the supply. If the price level is above the equilibrium level, people will want to hold more money than the Fed has created, so the price level must fall to balance supply and demand. If the price level is below the equilibrium level, people will want to hold less money than the Fed has created, and the price level must rise to balance supply and demand. At the equilibrium price level, the quantity of money that people want to hold exactly balances the quantity of money supplied by the Fed.

Figure 1 illustrates these ideas. The horizontal axis of this graph shows the quantity of money. The left vertical axis shows the value of money \(1/P\), and the right vertical axis shows the price level \(P\). Notice that the price-level axis on the right is inverted: A low price level is shown near the top of this axis, and a high price level is shown near the bottom. This inverted axis illustrates that when the value of money is high (as shown near the top of the left axis), the price level is low (as shown near the top of the right axis).

The two curves in this figure are the supply and demand curves for money. The supply curve is vertical because the Fed has fixed the quantity of money available. The demand curve for money is downward sloping, indicating that when the value of money is low (and the price level is high), people demand a larger quantity of it to buy goods and services. At the equilibrium, shown in the figure as point A, the quantity of money demanded balances the quantity of money supplied. This equilibrium of money supply and money demand determines the value of money and the price level.

**The Effects of a Monetary Injection**

Let’s now consider the effects of a change in monetary policy. To do so, imagine that the economy is in equilibrium and then, suddenly, the Fed doubles the supply of money by printing some dollar bills and dropping them around the country from helicopters. (Or less dramatically and more realistically, the Fed could inject money into the economy by buying some government bonds from the public in open-market operations.) What happens after such a monetary injection? How does the new equilibrium compare to the old one?
The horizontal axis shows the quantity of money. The left vertical axis shows the value of money, and the right vertical axis shows the price level. The supply curve for money is vertical because the quantity of money supplied is fixed by the Fed. The demand curve for money is downward sloping because people want to hold a larger quantity of money when each dollar buys less. At the equilibrium, point A, the value of money (on the left axis) and the price level (on the right axis) have adjusted to bring the quantity of money supplied and the quantity of money demanded into balance.

Figure 2 shows what happens. The monetary injection shifts the supply curve to the right from $MS_1$ to $MS_2$, and the equilibrium moves from point A to point B. As a result, the value of money (shown on the left axis) decreases from $\frac{1}{2}$ to $\frac{1}{4}$, and the equilibrium price level (shown on the right axis) increases from 2 to 4. In other words, when an increase in the money supply makes dollars more plentiful, the result is an increase in the price level that makes each dollar less valuable.

This explanation of how the price level is determined and why it might change over time is called the **quantity theory of money**. According to the quantity theory, the quantity of money available in an economy determines the value of money, and growth in the quantity of money is the primary cause of inflation. As economist Milton Friedman once put it, “Inflation is always and everywhere a monetary phenomenon.”

**A BRIEF LOOK AT THE ADJUSTMENT PROCESS**

So far, we have compared the old equilibrium and the new equilibrium after an injection of money. How does the economy move from the old to the new equilibrium? A complete answer to this question requires an understanding of short-run fluctuations in the economy, which we examine later in this book. Here, we briefly consider the adjustment process that occurs after a change in the money supply.

The immediate effect of a monetary injection is to create an excess supply of money. Before the injection, the economy was in equilibrium (point A in Figure 2).
At the prevailing price level, people had exactly as much money as they wanted. But after the helicopters drop the new money and people pick it up off the streets, people have more dollars in their wallets than they want. At the prevailing price level, the quantity of money supplied now exceeds the quantity demanded. People try to get rid of this excess supply of money in various ways. They might use it to buy goods and services. Or they might use this excess money to make loans to others by buying bonds or by depositing the money in a bank savings account. These loans allow other people to buy goods and services. In either case, the injection of money increases the demand for goods and services.

The economy’s ability to supply goods and services, however, has not changed. As we saw in the chapter on production and growth, the economy’s output of goods and services is determined by the available labor, physical capital, human capital, natural resources, and technological knowledge. None of these is altered by the injection of money.

Thus, the greater demand for goods and services causes the prices of goods and services to increase. The increase in the price level, in turn, increases the quantity of money demanded because people are using more dollars for every transaction. Eventually, the economy reaches a new equilibrium (point B in Figure 2) at which the quantity of money demanded again equals the quantity of money supplied. In this way, the overall price level for goods and services adjusts to bring money supply and money demand into balance.
We have seen how changes in the money supply lead to changes in the average level of prices of goods and services. How do monetary changes affect other economic variables, such as production, employment, real wages, and real interest rates? This question has long intrigued economists, including David Hume in the 18th century.

Hume and his contemporaries suggested that economic variables should be divided into two groups. The first group consists of nominal variables—variables measured in monetary units. The second group consists of real variables—variables measured in physical units. For example, the income of corn farmers is a nominal variable because it is measured in dollars, whereas the quantity of corn they produce is a real variable because it is measured in bushels. Nominal GDP is a nominal variable because it measures the dollar value of the economy’s output of goods and services; real GDP is a real variable because it measures the total quantity of goods and services produced and is not influenced by the current prices of those goods and services. The separation of real and nominal variables is now called the classical dichotomy. (A dichotomy is a division into two groups, and classical refers to the earlier economic thinkers.)

Application of the classical dichotomy is tricky when we turn to prices. Most prices are quoted in units of money and, therefore, are nominal variables. When we say that the price of corn is $2 a bushel or that the price of wheat is $1 a bushel, both prices are nominal variables. But what about a relative price—the price of one thing compared to another? In our example, we could say that the price of a bushel of corn is 2 bushels of wheat. This relative price is not measured in terms of money. When comparing the prices of any two goods, the dollar signs cancel, and the resulting number is measured in physical units. Thus, while dollar prices are nominal variables, relative prices are real variables.

This lesson has many applications. For instance, the real wage (the dollar wage adjusted for inflation) is a real variable because it measures the rate at which people exchange goods and services for a unit of labor. Similarly, the real interest rate (the nominal interest rate adjusted for inflation) is a real variable because it measures the rate at which people exchange goods and services today for goods and services in the future.

Why separate variables into these groups? The classical dichotomy is useful because different forces influence real and nominal variables. According to classical analysis, nominal variables are influenced by developments in the economy’s monetary system, whereas money is largely irrelevant for explaining real variables.

This idea was implicit in our discussion of the real economy in the long run. In previous chapters, we examined how real GDP, saving, investment, real interest rates, and unemployment are determined without mentioning the existence of money. In that analysis, the economy’s production of goods and services depends on productivity and factor supplies, the real interest rate balances the supply and demand for loanable funds, the real wage balances the supply and demand for labor, and unemployment results when the real wage is for some reason kept above its equilibrium level. These conclusions have nothing to do with the quantity of money supplied.
Changes in the supply of money, according to classical analysis, affect nominal variables but not real ones. When the central bank doubles the money supply, the price level doubles, the dollar wage doubles, and all other dollar values double. Real variables, such as production, employment, real wages, and real interest rates, are unchanged. The irrelevance of monetary changes for real variables is called monetary neutrality.

An analogy helps explain monetary neutrality. As the unit of account, money is the yardstick we use to measure economic transactions. When a central bank doubles the money supply, all prices double, and the value of the unit of account falls by half. A similar change would occur if the government were to reduce the length of the yard from 36 to 18 inches: With the new unit of measurement, all measured distances (nominal variables) would double, but the actual distances (real variables) would remain the same. The dollar, like the yard, is merely a unit of measurement, so a change in its value should not have real effects.

Is monetary neutrality realistic? Not completely. A change in the length of the yard from 36 to 18 inches would not matter in the long run, but in the short run, it would lead to confusion and mistakes. Similarly, most economists today believe that over short periods of time—within the span of a year or two—monetary changes affect real variables. Hume himself also doubted that monetary neutrality would apply in the short run. (We will study short-run nonneutrality later in the book, and this topic will help explain why the Fed changes the money supply over time.)

Yet classical analysis is right about the economy in the long run. Over the course of a decade, monetary changes have significant effects on nominal variables (such as the price level) but only negligible effects on real variables (such as real GDP). When studying long-run changes in the economy, the neutrality of money offers a good description of how the world works.

**Velocity and the Quantity Equation**

We can obtain another perspective on the quantity theory of money by considering the following question: How many times per year is the typical dollar bill used to pay for a newly produced good or service? The answer to this question is given by a variable called the velocity of money. In physics, the term velocity refers to the speed at which an object travels. In economics, the velocity of money refers to the speed at which the typical dollar bill travels around the economy from wallet to wallet.

To calculate the velocity of money, we divide the nominal value of output (nominal GDP) by the quantity of money. If \( P \) is the price level (the GDP deflator), \( Y \) the quantity of output (real GDP), and \( M \) the quantity of money, then velocity is

\[
V = \frac{(P \times Y)}{M}.
\]

To see why this makes sense, imagine a simple economy that produces only pizza. Suppose that the economy produces 100 pizzas in a year, that a pizza sells for $10, and that the quantity of money in the economy is $50. Then the velocity of money is

\[
V = \frac{($10 \times 100)}{50} = 20.
\]
In this economy, people spend a total of $1,000 per year on pizza. For this $1,000 of spending to take place with only $50 of money, each dollar bill must change hands on average 20 times per year.

With slight algebraic rearrangement, this equation can be rewritten as

\[ M \times V = P \times Y. \]

This equation states that the quantity of money \((M)\) times the velocity of money \((V)\) equals the price of output \((P)\) times the amount of output \((Y)\). It is called the quantity equation because it relates the quantity of money \((M)\) to the nominal value of output \((P \times Y)\). The quantity equation shows that an increase in the quantity of money in an economy must be reflected in one of the other three variables: The price level must rise, the quantity of output must rise, or the velocity of money must fall.

In many cases, it turns out that the velocity of money is relatively stable. For example, Figure 3 shows nominal GDP, the quantity of money (as measured by M2), and the velocity of money for the U.S. economy since 1960. During the period, the money supply and nominal GDP both increased more than twentyfold. By

---

**Figure 3**

Nominal GDP, the Quantity of Money, and the Velocity of Money

This figure shows the nominal value of output as measured by nominal GDP, the quantity of money as measured by M2, and the velocity of money as measured by their ratio. For comparability, all three series have been scaled to equal 100 in 1960. Notice that nominal GDP and the quantity of money have grown dramatically over this period, while velocity has been relatively stable.

*Source: U.S. Department of Commerce; Federal Reserve Board.*
contrast, the velocity of money, although not exactly constant, has not changed
dramatically. Thus, for some purposes, the assumption of constant velocity may
be a good approximation.

We now have all the elements necessary to explain the equilibrium price level
and inflation rate. Here they are:

1. The velocity of money is relatively stable over time.
2. Because velocity is stable, when the central bank changes the quantity of
money \((M)\), it causes proportionate changes in the nominal value of output
\((P \times Y)\).
3. The economy’s output of goods and services \((Y)\) is primarily determined
by factor supplies (labor, physical capital, human capital, and natural
resources) and the available production technology. In particular, because
money is neutral, money does not affect output.
4. With output \((Y)\) determined by factor supplies and technology, when the
central bank alters the money supply \((M)\) and induces proportional changes
in the nominal value of output \((P \times Y)\), these changes are reflected in
changes in the price level \((P)\).
5. Therefore, when the central bank increases the money supply rapidly, the
result is a high rate of inflation.

These five steps are the essence of the quantity theory of money.

**CASE STUDY**

**MONEY AND PRICES DURING FOUR HYPERINFLATIONS**

Although earthquakes can wreak havoc on a society, they have the beneficial by-
product of providing much useful data for seismologists. These data can shed
light on alternative theories and, thereby, help society predict and deal with future
threats. Similarly, hyperinflations offer monetary economists a natural experiment
they can use to study the effects of money on the economy.

Hyperinflations are interesting in part because the changes in the money supply
and price level are so large. Indeed, hyperinflation is generally defined as inflation
that exceeds 50 percent *per month*. This means that the price level increases more
than a hundredfold over the course of a year.

The data on hyperinflation show a clear link between the quantity of money
and the price level. Figure 4 graphs data from four classic hyperinflations that
occurred during the 1920s in Austria, Hungary, Germany, and Poland. Each graph
shows the quantity of money in the economy and an index of the price level. The
slope of the money line represents the rate at which the quantity of money was
growing, and the slope of the price line represents the inflation rate. The steeper
the lines, the higher the rates of money growth or inflation.

Notice that in each graph the quantity of money and the price level are almost
parallel. In each instance, growth in the quantity of money is moderate at first and
so is inflation. But over time, the quantity of money in the economy starts grow-
ing faster and faster. At about the same time, inflation also takes off. Then when
the quantity of money stabilizes, the price level stabilizes as well. These episodes
illustrate well one of the *Ten Principles of Economics*: Prices rise when the govern-
ment prints too much money.
This figure shows the quantity of money and the price level during four hyperinflations. (Note that these variables are graphed on logarithmic scales. This means that equal vertical distances on the graph represent equal percentage changes in the variable.) In each case, the quantity of money and the price level move closely together. The strong association between these two variables is consistent with the quantity theory of money, which states that growth in the money supply is the primary cause of inflation.


---

**The Inflation Tax**

If inflation is so easy to explain, why do countries experience hyperinflation? That is, why do the central banks of these countries choose to print so much money that its value is certain to fall rapidly over time?

The answer is that the governments of these countries are using money creation as a way to pay for their spending. When the government wants to build roads, pay salaries to its soldiers, or give transfer payments to the poor or elderly, it first has to raise the necessary funds. Normally, the government does this by levying taxes, such as income and sales taxes, and by borrowing from the public by selling government bonds. Yet the government can also pay for spending simply by printing the money it needs.
When the government raises revenue by printing money, it is said to levy an **inflation tax**. The inflation tax is not exactly like other taxes, however, because no one receives a bill from the government for this tax. Instead, the inflation tax is subtler. When the government prints money, the price level rises, and the dollars in your wallet are less valuable. Thus, the inflation tax is like a tax on everyone who holds money.

The importance of the inflation tax varies from country to country and over time. In the United States in recent years, the inflation tax has been a trivial source of revenue: It has accounted for less than 3 percent of government revenue. During the 1770s, however, the Continental Congress of the fledgling United States relied heavily on the inflation tax to pay for military spending. Because the new government had a limited ability to raise funds through regular taxes or borrowing, printing dollars was the easiest way to pay the American soldiers. As the quantity theory predicts, the result was a high rate of inflation: Prices measured in terms of the continental dollar rose more than a hundredfold over a few years.

Almost all hyperinflations follow the same pattern as the hyperinflation during the American Revolution. The government has high spending, inadequate tax revenue, and limited ability to borrow. As a result, it turns to the printing press to pay for its spending. The massive increases in the quantity of money lead to massive inflation. The inflation ends when the government institutes fiscal reforms—such as cuts in government spending—that eliminate the need for the inflation tax.

### THE FISHER EFFECT

According to the principle of monetary neutrality, an increase in the rate of money growth raises the rate of inflation but does not affect any real variable. An important application of this principle concerns the effect of money on interest rates. Interest rates are important variables for macroeconomists to understand because they link the economy of the present and the economy of the future through their effects on saving and investment.

To understand the relationship between money, inflation, and interest rates, recall the distinction between the nominal interest rate and the real interest rate. The **nominal interest rate** is the interest rate you hear about at your bank. If you have a savings account, for instance, the nominal interest rate tells you how fast the number of dollars in your account will rise over time. The **real interest rate** corrects the nominal interest rate for the effect of inflation to tell you how fast the purchasing power of your savings account will rise over time. The real interest rate is the nominal interest rate minus the inflation rate:

\[
\text{Real interest rate} = \text{Nominal interest rate} - \text{Inflation rate}.
\]

For example, if the bank posts a nominal interest rate of 7 percent per year and the inflation rate is 3 percent per year, then the real value of the deposits grows by 4 percent per year.

We can rewrite this equation to show that the nominal interest rate is the sum of the real interest rate and the inflation rate:

\[
\text{Nominal interest rate} = \text{Real interest rate} + \text{Inflation rate}.
\]
**Freeze on Wages Is Latest Step to Stanch Inflation in Zimbabwe**

By Michael Wines

Zimbabwe’s government slapped a six-month freeze on wages, rents and service fees on Friday, the latest step in what some analysts call an increasingly desperate campaign to sustain an economy gutted by hyperinflation.

Even as President Robert G. Mugabe declared the freeze, however, Zimbabwean newspapers suggested that the government’s two-month-old drive against inflation had backfired by drying up tax revenues needed to run the government.

The new freeze, announced in Friday’s editions of government-controlled newspapers, is intended to combat an annual inflation rate that the government says exceeds 7,600 percent, and private economists say is twice that. It bars businesses from indexing wages or fees to inflation, a method employed in many wage agreements.

All increases must now be approved by a government commission, the state-run Herald newspaper reported.

The freeze follows a decree issued in late June that forced merchants and wholesalers to reduce all prices by at least 50 percent. Shoppers stripped store shelves of clothes, meat and other basic goods after that decree, and producers have largely failed to ship new stock because goods now sell for less than it costs to make them.

Most commodities are now available only on the black market, where prices have continued to skyrocket. Moreover, as the last remaining stocks of goods trickle out of factory warehouses and onto the market, Zimbabwe could soon see the start of an inflationary spiral that would make today’s prices seem cheap, John Robertson, a Harare economist, said in an interview.

“It could go much higher—10 times as much for some things in the next couple of weeks, as goods cease to exist,” he said.

Mr. Robertson said idle producers had been forced to lay off workers to cut costs, cutting the government’s payroll tax receipts, and that sales-tax revenues were plummeting because stores had little to buy.

Harare’s Financial Gazette newspaper, which is controlled by the president of the government’s reserve bank, Gideon Gono, reported in this week’s edition that value-added tax receipts had dropped by up to 90 percent since the price-cutting campaign began.

The Zimbabwe Independent, one of the few newspapers not under government ownership, reported that the price cuts had cost the government 13 trillion Zimbabwe dollars in lost tax revenue. At current black market rates, that totals about $55 million—a vast sum for a government that is already technically bankrupt.

The government continues to function by printing money to pay its bills, but as the currency has dwindled in value, state workers have increasingly demanded regular raises. Zimbabwe’s 100,000 teachers, all government employees, have been threatening to strike if their pay is not increased.

The military, which is among Mr. Mugabe’s most reliable supporters, is also asking for wage increases for soldiers. A report issued this week by the Parliament’s defense and home affairs committee warned that the military was running out of money to pay foreign suppliers and maintain its infrastructure.

This way of looking at the nominal interest rate is useful because different economic forces determine each of the two terms on the right side of this equation. As we discussed earlier in the book, the supply and demand for loanable funds determine the real interest rate. And according to the quantity theory of money, growth in the money supply determines the inflation rate.

Let’s now consider how the growth in the money supply affects interest rates. In the long run over which money is neutral, a change in money growth should not affect the real interest rate. The real interest rate is, after all, a real variable. For the real interest rate not to be affected, the nominal interest rate must adjust one-for-one to changes in the inflation rate. Thus, when the Fed increases the rate of money growth, the long-run result is both a higher inflation rate and a higher nominal interest rate. This adjustment of the nominal interest rate to the inflation rate is called the Fisher effect, after economist Irving Fisher (1867–1947), who first studied it.

Keep in mind that our analysis of the Fisher effect has maintained a long-run perspective. The Fisher effect need not hold in the short run because inflation may be unanticipated. A nominal interest rate is a payment on a loan, and it is typically set when the loan is first made. If a jump in inflation catches the borrower and lender by surprise, the nominal interest rate they agreed on will fail to reflect the higher inflation. But if inflation remains high, people will eventually come to expect it, and loan agreements will reflect this expectation. To be precise, therefore, the Fisher effect states that the nominal interest rate adjusts to expected inflation. Expected inflation moves with actual inflation in the long run, but that is not necessarily true in the short run.

The Fisher effect is crucial for understanding changes over time in the nominal interest rate. Figure 5 shows the nominal interest rate and the inflation rate in the U.S. economy since 1960. The close association between these two variables is clear. The nominal interest rate rose from the early 1960s through the 1970s because inflation was also rising during this time. Similarly, the nominal interest rate fell from the early 1980s through the 1990s because the Fed got inflation under control.

**QUICK QUIZ** The government of a country increases the growth rate of the money supply from 5 percent per year to 50 percent per year. What happens to prices? What happens to nominal interest rates? Why might the government be doing this?

---

**THE COSTS OF INFLATION**

In the late 1970s, when the U.S. inflation rate reached about 10 percent per year, inflation dominated debates over economic policy. And even though inflation has been low over the past decade, it remains a closely watched macroeconomic variable. One study found that inflation is the economic term mentioned most often in U.S. newspapers (far ahead of second-place finisher unemployment and third-place finisher productivity).

Inflation is closely watched and widely discussed because it is thought to be a serious economic problem. But is that true? And if so, why?
A Fall in Purchasing Power? The Inflation Fallacy

If you ask the typical person why inflation is bad, he will tell you that the answer is obvious: Inflation robs him of the purchasing power of his hard-earned dollars. When prices rise, each dollar of income buys fewer goods and services. Thus, it might seem that inflation directly lowers living standards.

Yet further thought reveals a fallacy in this answer. When prices rise, buyers of goods and services pay more for what they buy. At the same time, however, sellers of goods and services get more for what they sell. Because most people earn their incomes by selling their services, such as their labor, inflation in incomes goes hand in hand with inflation in prices. Thus, inflation does not in itself reduce people’s real purchasing power.

People believe the inflation fallacy because they do not appreciate the principle of monetary neutrality. A worker who receives an annual raise of 10 percent tends to view that raise as a reward for her own talent and effort. When an inflation rate of 6 percent reduces the real value of that raise to only 4 percent, the worker might feel that she has been cheated of what is rightfully her due. In fact, as we discussed in the chapter on production and growth, real incomes are determined by real variables, such as physical capital, human capital, natural resources, and the available production technology. Nominal incomes are determined by those factors and the overall price level. If the Fed were to lower the inflation rate from 6 percent to zero, our worker’s annual raise would fall from 10 percent to
4 percent. She might feel less robbed by inflation, but her real income would not rise more quickly.

If nominal incomes tend to keep pace with rising prices, why then is inflation a problem? It turns out that there is no single answer to this question. Instead, economists have identified several costs of inflation. Each of these costs shows some way in which persistent growth in the money supply does, in fact, have some effect on real variables.

**Shoeleather Costs**

As we have discussed, inflation is like a tax on the holders of money. The tax itself is not a cost to society: It is only a transfer of resources from households to the government. Yet most taxes give people an incentive to alter their behavior to avoid paying the tax, and this distortion of incentives causes deadweight losses for society as a whole. Like other taxes, the inflation tax also causes deadweight losses because people waste scarce resources trying to avoid it.

How can a person avoid paying the inflation tax? Because inflation erodes the real value of the money in your wallet, you can avoid the inflation tax by holding less money. One way to do this is to go to the bank more often. For example, rather than withdrawing $200 every four weeks, you might withdraw $50 once a week. By making more frequent trips to the bank, you can keep more of your wealth in your interest-bearing savings account and less in your wallet, where inflation erodes its value.

The cost of reducing your money holdings is called the **shoeleather cost** of inflation because making more frequent trips to the bank causes your shoes to wear out more quickly. Of course, this term is not to be taken literally: The actual cost of reducing your money holdings is not the wear and tear on your shoes but the time and convenience you must sacrifice to keep less money on hand than you would if there were no inflation.

The shoeleather costs of inflation may seem trivial. And in fact, they are in the U.S. economy, which has had only moderate inflation in recent years. But this cost is magnified in countries experiencing hyperinflation. Here is a description of one person’s experience in Bolivia during its hyperinflation (as reported in the August 13, 1985, issue of *The Wall Street Journal*):

When Edgar Miranda gets his monthly teacher’s pay of 25 million pesos, he hasn’t a moment to lose. Every hour, pesos drop in value. So, while his wife rushes to market to lay in a month’s supply of rice and noodles, he is off with the rest of the pesos to change them into black-market dollars.

Mr. Miranda is practicing the First Rule of Survival amid the most out-of-control inflation in the world today. Bolivia is a case study of how runaway inflation undermines a society. Price increases are so huge that the figures build up almost beyond comprehension. In one six-month period, for example, prices soared at an annual rate of 38,000 percent. By official count, however, last year’s inflation reached 2,000 percent, and this year’s is expected to hit 8,000 percent—though other estimates range many times higher. In any event, Bolivia’s rate dwarfs Israel’s 370 percent and Argentina’s 1,100 percent—two other cases of severe inflation.

It is easier to comprehend what happens to the thirty-eight-year-old Mr. Miranda’s pay if he doesn’t quickly change it into dollars. The day he was paid
25 million pesos, a dollar cost 500,000 pesos. So he received $50. Just days later, with the rate at 900,000 pesos, he would have received $27.

As this story shows, the shoeleather costs of inflation can be substantial. With the high inflation rate, Mr. Miranda does not have the luxury of holding the local money as a store of value. Instead, he is forced to convert his pesos quickly into goods or into U.S. dollars, which offer a more stable store of value. The time and effort that Mr. Miranda expends to reduce his money holdings are a waste of resources. If the monetary authority pursued a low-inflation policy, Mr. Miranda would be happy to hold pesos, and he could put his time and effort to more productive use. In fact, shortly after this article was written, the Bolivian inflation rate was reduced substantially with a more restrictive monetary policy.

**Menu Costs**

Most firms do not change the prices of their products every day. Instead, firms often announce prices and leave them unchanged for weeks, months, or even years. One survey found that the typical U.S. firm changes its prices about once a year.

Firms change prices infrequently because there are costs of changing prices. Costs of price adjustment are called **menu costs**, a term derived from a restaurant’s cost of printing a new menu. Menu costs include the cost of deciding on new prices, the cost of printing new price lists and catalogs, the cost of sending these new price lists and catalogs to dealers and customers, the cost of advertising the new prices, and even the cost of dealing with customer annoyance over price changes.

Inflation increases the menu costs that firms must bear. In the current U.S. economy, with its low inflation rate, annual price adjustment is an appropriate business strategy for many firms. But when high inflation makes firms’ costs rise rapidly, annual price adjustment is impractical. During hyperinflations, for example, firms must change their prices daily or even more often just to keep up with all the other prices in the economy.

**Relative-Price Variability and the Misallocation of Resources**

Suppose that the Eatabit Eatery prints a new menu with new prices every January and then leaves its prices unchanged for the rest of the year. If there is no inflation, Eatabit’s relative prices—the prices of its meals compared to other prices in the economy—would be constant over the course of the year. By contrast, if the inflation rate is 12 percent per year, Eatabit’s relative prices will automatically fall by 1 percent each month. The restaurant’s relative prices will be high in the early months of the year, just after it has printed a new menu, and low in the later months. And the higher the inflation rate, the greater is this automatic variability. Thus, because prices change only once in a while, inflation causes relative prices to vary more than they otherwise would.

Why does this matter? The reason is that market economies rely on relative prices to allocate scarce resources. Consumers decide what to buy by comparing the quality and prices of various goods and services. Through these decisions, they determine how the scarce factors of production are allocated among...
industries and firms. When inflation distorts relative prices, consumer decisions are distorted, and markets are less able to allocate resources to their best use.

**Inflation-Induced Tax Distortions**

Almost all taxes distort incentives, cause people to alter their behavior, and lead to a less efficient allocation of the economy’s resources. Many taxes, however, become even more problematic in the presence of inflation. The reason is that lawmakers often fail to take inflation into account when writing the tax laws. Economists who have studied the tax code conclude that inflation tends to raise the tax burden on income earned from savings.

One example of how inflation discourages saving is the tax treatment of *capital gains*—the profits made by selling an asset for more than its purchase price. Suppose that in 1980 you used some of your savings to buy stock in Microsoft Corporation for $10 and that in 2008 you sold the stock for $50. According to the tax law, you have earned a capital gain of $40, which you must include in your income when computing how much income tax you owe. But suppose the overall price level doubled from 1980 to 2008. In this case, the $10 you invested in 1980 is equivalent (in terms of purchasing power) to $20 in 2008. When you sell your stock for $50, you have a real gain (an increase in purchasing power) of only $30. The tax code, however, does not take account of inflation and assesses you a tax on a gain of $40. Thus, inflation exaggerates the size of capital gains and inadvertently increases the tax burden on this type of income.

Another example is the tax treatment of interest income. The income tax treats the *nominal* interest earned on savings as income, even though part of the nominal interest rate merely compensates for inflation. To see the effects of this policy, consider the numerical example in Table 1. The table compares two economies, both of which tax interest income at a rate of 25 percent. In Economy A, inflation is zero, and the nominal and real interest rates are both 4 percent. In Economy B, the real interest rate is again 4 percent, but the inflation

<table>
<thead>
<tr>
<th>Table</th>
<th>How Inflation Raises the Tax Burden on Saving</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Economy A (price stability)</td>
</tr>
<tr>
<td>Real interest rate</td>
<td>4%</td>
</tr>
<tr>
<td>Inflation rate</td>
<td>0</td>
</tr>
<tr>
<td>Nominal interest rate</td>
<td>4</td>
</tr>
<tr>
<td>Reduced interest due to 25 percent tax (.25 × nominal interest rate)</td>
<td>1</td>
</tr>
<tr>
<td>After-tax nominal interest rate (.75 × nominal interest rate)</td>
<td>3</td>
</tr>
<tr>
<td>After-tax real interest rate (after-tax nominal interest rate – inflation rate)</td>
<td>3</td>
</tr>
</tbody>
</table>
rate is 8 percent. As a result of the Fisher effect, the nominal interest rate is 12 percent. Because the income tax treats this entire 12 percent interest as income, the government takes 25 percent of it, leaving an after-tax nominal interest rate of only 9 percent and an after-tax real interest rate of only 1 percent. In this case, the 25 percent tax on interest income reduces the real interest rate from 4 percent to 1 percent. Because the after-tax real interest rate provides the incentive to save, saving is much less attractive in the economy with inflation (Economy B) than in the economy with stable prices (Economy A).

The taxes on nominal capital gains and on nominal interest income are two examples of how the tax code interacts with inflation. There are many others. Because of these inflation-induced tax changes, higher inflation tends to discourage people from saving. Recall that the economy’s saving provides the resources for investment, which in turn is a key ingredient to long-run economic growth. Thus, when inflation raises the tax burden on saving, it tends to depress the economy’s long-run growth rate. There is, however, no consensus among economists about the size of this effect.

One solution to this problem, other than eliminating inflation, is to index the tax system. That is, the tax laws could be rewritten to take account of the effects of inflation. In the case of capital gains, for example, the tax code could adjust the purchase price using a price index and assess the tax only on the real gain. In the case of interest income, the government could tax only real interest income by excluding that portion of the interest income that merely compensates for inflation. To some extent, the tax laws have moved in the direction of indexation. For example, the income levels at which income tax rates change are adjusted automatically each year based on changes in the consumer price index. Yet many other aspects of the tax laws—such as the tax treatment of capital gains and interest income—are not indexed.

In an ideal world, the tax laws would be written so that inflation would not alter anyone’s real tax liability. In the world in which we live, however, tax laws are far from perfect. More complete indexation would probably be desirable, but it would further complicate a tax code that many people already consider too complex.

**Confusion and Inconvenience**

Imagine that we took a poll and asked people the following question: “This year the yard is 36 inches. How long do you think it should be next year?” Assuming we could get people to take us seriously, they would tell us that the yard should stay the same length—36 inches. Anything else would just complicate life needlessly.

What does this finding have to do with inflation? Recall that money, as the economy’s unit of account, is what we use to quote prices and record debts. In other words, money is the yardstick with which we measure economic transactions. The job of the Federal Reserve is a bit like the job of the Bureau of Standards—to ensure the reliability of a commonly used unit of measurement. When the Fed increases the money supply and creates inflation, it erodes the real value of the unit of account.

It is difficult to judge the costs of the confusion and inconvenience that arise from inflation. Earlier, we discussed how the tax code incorrectly measures real incomes in the presence of inflation. Similarly, accountants incorrectly measure firms’ earnings when prices are rising over time. Because inflation causes dollars
at different times to have different real values, computing a firm’s profit—the difference between its revenue and costs—is more complicated in an economy with inflation. Therefore, to some extent, inflation makes investors less able to sort successful from unsuccessful firms, which in turn impedes financial markets in their role of allocating the economy’s saving to alternative types of investment.

**A Special Cost of Unexpected Inflation: Arbitrary Redistributions of Wealth**

So far, the costs of inflation we have discussed occur even if inflation is steady and predictable. Inflation has an additional cost, however, when it comes as a surprise. Unexpected inflation redistributes wealth among the population in a way that has nothing to do with either merit or need. These redistributions occur because many loans in the economy are specified in terms of the unit of account—money.

Consider an example. Suppose that Sam Student takes out a $20,000 loan at a 7 percent interest rate from Bigbank to attend college. In 10 years, the loan will come due. After his debt has compounded for 10 years at 7 percent, Sam will owe Bigbank $40,000. The real value of this debt will depend on inflation over the decade. If Sam is lucky, the economy will have a hyperinflation. In this case, wages and prices will rise so high that Sam will be able to pay the $40,000 debt out of pocket change. By contrast, if the economy goes through a major deflation, then wages and prices will fall, and Sam will find the $40,000 debt a greater burden than he anticipated.

This example shows that unexpected changes in prices redistribute wealth among debtors and creditors. A hyperinflation enriches Sam at the expense of Bigbank because it diminishes the real value of the debt; Sam can repay the loan in less valuable dollars than he anticipated. Deflation enriches Bigbank at Sam’s expense because it increases the real value of the debt; in this case, Sam has to repay the loan in more valuable dollars than he anticipated. If inflation were predictable, then Bigbank and Sam could take inflation into account when setting the nominal interest rate. (Recall the Fisher effect.) But if inflation is hard to predict, it imposes risk on Sam and Bigbank that both would prefer to avoid.

This cost of unexpected inflation is important to consider together with another fact: Inflation is especially volatile and uncertain when the average rate of inflation is high. This is seen most simply by examining the experience of different countries. Countries with low average inflation, such as Germany in the late 20th century, tend to have stable inflation. Countries with high average inflation, such as many countries in Latin America, tend to have unstable inflation. There are no known examples of economies with high, stable inflation. This relationship between the level and volatility of inflation points to another cost of inflation. If a country pursues a high-inflation monetary policy, it will have to bear not only the costs of high expected inflation but also the arbitrary redistributions of wealth associated with unexpected inflation.

**Case Study: The Wizard of Oz and the Free-Silver Debate**

As a child, you probably saw the movie *The Wizard of Oz*, based on a children’s book written in 1900. The movie and book tell the story of a young girl, Dorothy, who finds herself lost in a strange land far from home. You probably did not
know, however, that the story is actually an allegory about U.S. monetary policy in the late 19th century.

From 1880 to 1896, the price level in the U.S. economy fell by 23 percent. Because this event was unanticipated, it led to a major redistribution of wealth. Most farmers in the western part of the country were debtors. Their creditors were the bankers in the east. When the price level fell, it caused the real value of these debts to rise, which enriched the banks at the expense of the farmers.

According to Populist politicians of the time, the solution to the farmers’ problem was the free coinage of silver. During this period, the United States was operating with a gold standard. The quantity of gold determined the money supply and, thereby, the price level. The free-silver advocates wanted silver, as well as gold, to be used as money. If adopted, this proposal would have increased the money supply, pushed up the price level, and reduced the real burden of the farmers’ debts.

The debate over silver was heated, and it was central to the politics of the 1890s. A common election slogan of the Populists was “We Are Mortgaged. All but Our Votes.” One prominent advocate of free silver was William Jennings Bryan, the Democratic nominee for president in 1896. He is remembered in part for a speech at the Democratic Party’s nominating convention in which he said, “You shall not press down upon the brow of labor this crown of thorns. You shall not crucify mankind upon a cross of gold.” Rarely since then have politicians waxed so poetic about alternative approaches to monetary policy. Nonetheless, Bryan lost the election to Republican William McKinley, and the United States remained on the gold standard.

L. Frank Baum, author of the book *The Wonderful Wizard of Oz*, was a midwestern journalist. When he sat down to write a story for children, he made the characters represent protagonists in the major political battle of his time. Here is how economic historian Hugh Rockoff, writing in the *Journal of Political Economy* in 1990, interprets the story:

- **Dorothy:** Traditional American values
- **Toto:** Prohibitionist party, also called the Teetotalers
- **Scarecrow:** Farmers
- **Tin Woodman:** Industrial workers
- **Cowardly Lion:** William Jennings Bryan
- **Munchkins:** Citizens of the East
- **Wicked Witch of the East:** Grover Cleveland
- **Wicked Witch of the West:** William McKinley
- **Wizard:** Marcus Alonzo Hanna, chairman of the Republican Party
- **Oz:** Abbreviation for ounce of gold
- **Yellow Brick Road:** Gold standard

In the end of Baum’s story, Dorothy does find her way home, but it is not by just following the yellow brick road. After a long and perilous journey, she learns that the wizard is incapable of helping her or her friends. Instead, Dorothy finally discovers the magical power of her silver slippers. (When the book was made into a movie in 1939, Dorothy’s slippers were changed from silver to ruby. The Hollywood filmmakers were more interested in showing off the new technology of Technicolor than in telling a story about 19th-century monetary policy.)
Although the Populists lost the debate over the free coinage of silver, they did eventually get the monetary expansion and inflation that they wanted. In 1898, prospectors discovered gold near the Klondike River in the Canadian Yukon. Increased supplies of gold also arrived from the mines of South Africa. As a result, the money supply and the price level started to rise in the United States and other countries operating on the gold standard. Within 15 years, prices in the United States were back to the levels that had prevailed in the 1880s, and farmers were better able to handle their debts.

**QUICK QUIZ** List and describe six costs of inflation.

**CONCLUSION**

This chapter discussed the causes and costs of inflation. The primary cause of inflation is simply growth in the quantity of money. When the central bank creates money in large quantities, the value of money falls quickly. To maintain stable prices, the central bank must maintain strict control over the money supply.

The costs of inflation are subtler. They include shoeleather costs, menu costs, increased variability of relative prices, unintended changes in tax liabilities, confusion and inconvenience, and arbitrary redistributions of wealth. Are these costs, in total, large or small? All economists agree that they become huge during hyper-inflation. But their size for moderate inflation—when prices rise by less than 10 percent per year—is more open to debate.

Although this chapter presented many of the most important lessons about inflation, the discussion is incomplete. When the central bank reduces the rate of money growth, prices rise less rapidly, as the quantity theory suggests. Yet as the economy makes the transition to this lower inflation rate, the change in monetary policy will have disruptive effects on production and employment. That is, even though monetary policy is neutral in the long run, it has profound effects on real variables in the short run. Later in this book we will examine the reasons for short-run monetary nonneutrality to enhance our understanding of the causes and costs of inflation.
• The overall level of prices in an economy adjusts to bring money supply and money demand into balance. When the central bank increases the supply of money, it causes the price level to rise. Persistent growth in the quantity of money supplied leads to continuing inflation.
• The principle of monetary neutrality asserts that changes in the quantity of money influence nominal variables but not real variables. Most economists believe that monetary neutrality approximately describes the behavior of the economy in the long run.
• A government can pay for some of its spending simply by printing money. When countries rely heavily on this “inflation tax,” the result is hyperinflation.
• One application of the principle of monetary neutrality is the Fisher effect. According to the Fisher effect, when the inflation rate rises, the nominal interest rate rises by the same amount so that the real interest rate remains the same.
• Many people think that inflation makes them poorer because it raises the cost of what they buy. This view is a fallacy, however, because inflation also raises nominal incomes.
• Economists have identified six costs of inflation: shoeleather costs associated with reduced money holdings, menu costs associated with more frequent adjustment of prices, increased variability of relative prices, unintended changes in tax liabilities due to nonindexation of the tax code, confusion and inconvenience resulting from a changing unit of account, and arbitrary redistributions of wealth between debtors and creditors. Many of these costs are large during hyperinflation, but the size of these costs for moderate inflation is less clear.

Questions for Review

1. Explain how an increase in the price level affects the real value of money.
2. According to the quantity theory of money, what is the effect of an increase in the quantity of money?
3. Explain the difference between nominal and real variables and give two examples of each. According to the principle of monetary neutrality, which variables are affected by changes in the quantity of money?
4. In what sense is inflation like a tax? How does thinking about inflation as a tax help explain hyperinflation?
5. According to the Fisher effect, how does an increase in the inflation rate affect the real interest rate and the nominal interest rate?
6. What are the costs of inflation? Which of these costs do you think are most important for the U.S. economy?
7. If inflation is less than expected, who benefits—debtors or creditors? Explain.
1. Suppose that this year’s money supply is $500 billion, nominal GDP is $10 trillion, and real GDP is $5 trillion.
   a. What is the price level? What is the velocity of money?
   b. Suppose that velocity is constant and the economy’s output of goods and services rises by 5 percent each year. What will happen to nominal GDP and the price level next year if the Fed keeps the money supply constant?
   c. What money supply should the Fed set next year if it wants to keep the price level stable?
   d. What money supply should the Fed set next year if it wants inflation of 10 percent?

2. Suppose that changes in bank regulations expand the availability of credit cards so that people need to hold less cash.
   a. How does this event affect the demand for money?
   b. If the Fed does not respond to this event, what will happen to the price level?
   c. If the Fed wants to keep the price level stable, what should it do?

3. It is often suggested that the Federal Reserve try to achieve zero inflation. If we assume that velocity is constant, does this zero-inflation goal require that the rate of money growth equal zero? If yes, explain why. If no, explain what the rate of money growth should equal.

4. Suppose that a country’s inflation rate increases sharply. What happens to the inflation tax on the holders of money? Why is wealth that is held in savings accounts not subject to a change in the inflation tax? Can you think of any way holders of savings accounts are hurt by the increase in the inflation rate?

5. Hyperinflations are extremely rare in countries whose central banks are independent of the rest of the government. Why might this be so?

6. Let’s consider the effects of inflation in an economy composed of only two people: Bob, a bean farmer, and Rita, a rice farmer. Bob and Rita both always consume equal amounts of rice and beans. In 2008, the price of beans was $1, and the price of rice was $3.

   a. Suppose that in 2009 the price of beans was $2 and the price of rice was $6. What was inflation? Was Bob better off, worse off, or unaffected by the changes in prices? What about Rita?
   b. Now suppose that in 2009 the price of beans was $2 and the price of rice was $4. What was inflation? Was Bob better off, worse off, or unaffected by the changes in prices? What about Rita?
   c. Finally, suppose that in 2009 the price of beans was $2 and the price of rice was $1.50. What was inflation? Was Bob better off, worse off, or unaffected by the changes in prices? What about Rita?
   d. What matters more to Bob and Rita—the overall inflation rate or the relative price of rice and beans?

7. If the tax rate is 40 percent, compute the before-tax real interest rate and the after-tax real interest rate in each of the following cases.
   a. The nominal interest rate is 10 percent and the inflation rate is 5 percent.
   b. The nominal interest rate is 6 percent and the inflation rate is 2 percent.
   c. The nominal interest rate is 4 percent and the inflation rate is 1 percent.

8. What are your shoeleather costs of going to the bank? How might you measure these costs in dollars? How do you think the shoeleather costs of your college president differ from your own?

9. Recall that money serves three functions in the economy. What are those functions? How does inflation affect the ability of money to serve each of these functions?

10. Suppose that people expect inflation to equal 3 percent, but in fact, prices rise by 5 percent. Describe how this unexpectedly high inflation rate would help or hurt the following:
    a. the government
    b. a homeowner with a fixed-rate mortgage
    c. a union worker in the second year of a labor contract
    d. a college that has invested some of its endowment in government bonds
11. Explain one harm associated with unexpected inflation that is not associated with expected inflation. Then explain one harm associated with both expected and unexpected inflation.

12. Explain whether the following statements are true, false, or uncertain.
   
a. “Inflation hurts borrowers and helps lenders, because borrowers must pay a higher rate of interest.”

b. “If prices change in a way that leaves the overall price level unchanged, then no one is made better or worse off.”

c. “Inflation does not reduce the purchasing power of most workers.”
This page intentionally left blank
PART IX

Short-Run Economic Fluctuations
Economic activity fluctuates from year to year. In most years, the production of goods and services rises. Because of increases in the labor force, increases in the capital stock, and advances in technological knowledge, the economy can produce more and more over time. This growth allows everyone to enjoy a higher standard of living. On average over the past half century, the production of the U.S. economy as measured by real GDP has grown by about 3 percent per year.

In some years, however, the economy experiences contraction rather than growth. Firms find themselves unable to sell all the goods and services they have to offer, so they cut back on production. Workers are laid off, unemployment rises, and factories are left idle. With the economy producing fewer goods and services, real GDP and other measures of income fall. Such a period of falling incomes and rising unemployment is called a recession if it is relatively mild and a depression if it is more severe.

What causes short-run fluctuations in economic activity? What, if anything, can public policy do to prevent periods of falling incomes and rising unemployment? When recessions and depressions occur, how can policymakers reduce their length and severity? These are the questions we take up now.
The variables that we study are largely those we have already seen in previous chapters. They include GDP, unemployment, interest rates, and the price level. Also familiar are the policy instruments of government spending, taxes, and the money supply. What differs from our earlier analysis is the time horizon. So far, our goal has been to explain the behavior of these variables in the long run. Our goal now is to explain their short-run deviations from long-run trends. In other words, instead of focusing on the forces that explain economic growth from generation to generation, we are now interested in the forces that explain economic fluctuations from year to year.

There remains some debate among economists about how best to analyze short-run fluctuations, but most economists use the model of aggregate demand and aggregate supply. Learning how to use this model for analyzing the short-run effects of various events and policies is the primary task ahead. This chapter introduces the model’s two pieces: the aggregate-demand curve and the aggregate-supply curve. Before turning to the model, however, let’s look at some of the key facts that describe the ups and downs of the economy.

**THREE KEY FACTS ABOUT ECONOMIC FLUCTUATIONS**

Short-run fluctuations in economic activity occur in all countries throughout history. As a starting point for understanding these year-to-year fluctuations, let’s discuss some of their most important properties.

**FACT 1: ECONOMIC FLUCTUATIONS ARE IRREGULAR AND UNPREDICTABLE**

Fluctuations in the economy are often called the business cycle. As this term suggests, economic fluctuations correspond to changes in business conditions. When real GDP grows rapidly, business is good. During such periods of economic expansion, most firms find that customers are plentiful and that profits are growing. When real GDP falls during recessions, businesses have trouble. During such periods of economic contraction, most firms experience declining sales and dwindling profits.

The term business cycle is somewhat misleading because it suggests that economic fluctuations follow a regular, predictable pattern. In fact, economic fluctuations are not at all regular, and they are almost impossible to predict with much accuracy. Panel (a) of Figure 1 shows the real GDP of the U.S. economy since 1965. The shaded areas represent times of recession. As the figure shows, recessions do not come at regular intervals. Sometimes recessions are close together, such as the recessions of 1980 and 1982. Sometimes the economy goes many years without a recession. The longest period in U.S. history without a recession was the economic expansion from 1991 to 2001.

**FACT 2: MOST MACROECONOMIC QUANTITIES FLUCTUATE TOGETHER**

Real GDP is the variable most commonly used to monitor short-run changes in the economy because it is the most comprehensive measure of economic activity. Real GDP measures the value of all final goods and services produced within a
A Look at Short-Run Economic Fluctuations

This figure shows real GDP in panel (a), investment spending in panel (b), and unemployment in panel (c) for the U.S. economy using quarterly data since 1965. Recessions are shown as the shaded areas. Notice that real GDP and investment spending decline during recessions, while unemployment rises.

Source: U.S. Department of Commerce; U.S. Department of Labor.
It turns out, however, that for monitoring short-run fluctuations, it does not really matter which measure of economic activity one looks at. Most macroeconomic variables that measure some type of income, spending, or production fluctuate closely together. When real GDP falls in a recession, so do personal income, corporate profits, consumer spending, investment spending, industrial production, retail sales, home sales, auto sales, and so on. Because recessions are economy-wide phenomena, they show up in many sources of macroeconomic data.

Although many macroeconomic variables fluctuate together, they fluctuate by different amounts. In particular, as panel (b) of Figure 1 shows, investment spending varies greatly over the business cycle. Even though investment averages about one-seventh of GDP, declines in investment account for about two-thirds of the declines in GDP during recessions. In other words, when economic conditions deteriorate, much of the decline is attributable to reductions in spending on new factories, housing, and inventories.

**FACT 3: AS OUTPUT FALLS, UNEMPLOYMENT RISES**

Changes in the economy’s output of goods and services are strongly correlated with changes in the economy’s utilization of its labor force. In other words, when real GDP declines, the rate of unemployment rises. This fact is hardly surprising: When firms choose to produce a smaller quantity of goods and services, they lay off workers, expanding the pool of unemployed.

Panel (c) of Figure 1 shows the unemployment rate in the U.S. economy since 1965. Once again, the shaded areas in the figure indicate periods of recession. The figure shows clearly the impact of recessions on unemployment. In each of the recessions, the unemployment rate rises substantially. When the recession ends and real GDP starts to expand, the unemployment rate gradually declines. The unemployment rate never approaches zero; instead, it fluctuates around its natural rate of about 5 or 6 percent.

**QUICK QUIZ** List and discuss three key facts about economic fluctuations.

---

**EXPLAINING SHORT-RUN ECONOMIC FLUCTUATIONS**

Describing what happens to economies as they fluctuate over time is easy. Explaining what causes these fluctuations is more difficult. Indeed, compared to the topics we have studied in previous chapters, the theory of economic fluctuations remains controversial. In this and the next two chapters, we develop the model that most economists use to explain short-run fluctuations in economic activity.

**THE ASSUMPTIONS OF CLASSICAL ECONOMICS**

In previous chapters, we developed theories to explain what determines most important macroeconomic variables in the long run. Chapter 17 explained the level and growth of productivity and real GDP. Chapters 18 and 19 explained how the financial system works and how the real interest rate adjusts to balance saving and investment. Chapter 20 explained why there is always some unemployment in the
**Offbeat Indicators**

When the economy goes into a recession, many economic variables are affected. This article, written during the recession of 2001, gives some examples.

---

**Economic Numbers Befuddle Even the Best**

*By George Hager*

Economists pore over scores of numbers every week, trying to sense when the recession is over. But quirky indicators and gut instinct might be almost as helpful—maybe even more so.

Is your dentist busy? Dentists say people put off appointments when times turn tough, then reschedule when the economy improves.

How far away do you have to park when you go to the mall? Fewer shoppers equal more parking spaces.

If you drive to work, has your commute time gotten shorter or longer?

Economist Michael Evans says a colleague with a pipeline into the garbage business swears by his own homegrown Chicago Trash index. Collections plunged after the Sept. 11 terror attacks, rebounded in October but then fell off again in mid-November. "Trash is a pretty good indicator of what people are buying," says Evans, an economist with Evans Carroll & Associates. "They've got to throw out the wrappings." ... What's actually happening is often clear only in hindsight. That's one reason the National Bureau of Economic Research waited until November to declare that a recession began last March—and why it took them until December 1992 to declare that the last recession had ended more than a year earlier, in March 1991.

"The data can fail you," says Allen Sinai, chief economist for Decision Economics. Sinai cautions that if numbers appear to be going up (or down), it's best to wait to see what happens over two or three months before drawing a conclusion—something hair-trigger financial markets routinely don't do.

"We have lots of false predictions of recoveries by (stock) markets that don't happen," he says.

Like a lot of economists, Sinai leaves the numbers with informal observations. These days, he's paying particular attention to what business executives say at meetings and cocktail parties because their mood—and their plans for investing and hiring—are key to a comeback.

---

THE REALITY OF SHORT-RUN FLUCTUATIONS

Do these assumptions of classical macroeconomic theory apply to the world in which we live? The answer to this question is of central importance to understanding how the economy works. Most economists believe that classical theory describes the world in the long run but not in the short run.

All of this previous analysis was based on two related ideas: the classical dichotomy and monetary neutrality. Recall that the classical dichotomy is the separation of variables into real variables (those that measure quantities or relative prices) and nominal variables (those measured in terms of money). According to classical macroeconomic theory, changes in the money supply affect nominal variables but not real variables. As a result of this monetary neutrality, Chapters 17 through 20 were able to examine the determinants of real variables (real GDP, the real interest rate, and unemployment) without introducing nominal variables (the money supply and the price level).

In a sense, money does not matter in a classical world. If the quantity of money in the economy were to double, everything would cost twice as much, and everyone’s income would be twice as high. But so what? The change would be nominal (by the standard meaning of “nearly insignificant”). The things that people really care about—whether they have a job, how many goods and services they can afford, and so on—would be exactly the same.

This classical view is sometimes described by the saying, “Money is a veil.” That is, nominal variables may be the first things we see when we observe an economy because economic variables are often expressed in units of money. But what’s important are the real variables and the economic forces that determine them. According to classical theory, to understand these real variables, we need to look beneath the veil.

THE REALITY OF SHORT-RUN FLUCTUATIONS

Do these assumptions of classical macroeconomic theory apply to the world in which we live? The answer to this question is of central importance to understanding how the economy works. Most economists believe that classical theory describes the world in the long run but not in the short run.

Consider again the impact of money on the economy. Most economists believe that, beyond a period of several years, changes in the money supply affect prices and other nominal variables but do not affect real GDP, unemployment, or other real variables—just as classical theory says. When studying year-to-year changes in the economy, however, the assumption of monetary neutrality is no longer appropriate. In the short run, real and nominal variables are highly intertwined, and changes in the money supply can temporarily push real GDP away from its long-run trend.

Even the classical economists themselves, such as David Hume, realized that classical economic theory did not hold in the short run. From his vantage point in 18th-century England, Hume observed that when the money supply expanded after gold discoveries, it took some time for prices to rise, and in the meantime, the economy enjoyed higher employment and production.

To understand how the economy works in the short run, we need a new model. This new model can be built using many of the tools we developed in previous chapters, but it must abandon the classical dichotomy and the neutrality of money. We can no longer separate our analysis of real variables such as output and employment from our analysis of nominal variables such as money and the price level. Our new model focuses on how real and nominal variables interact.
The Model of Aggregate Demand and Aggregate Supply

Our model of short-run economic fluctuations focuses on the behavior of two variables. The first variable is the economy’s output of goods and services, as measured by real GDP. The second is the average level of prices, as measured by the CPI or the GDP deflator. Notice that output is a real variable, whereas the price level is a nominal variable. By focusing on the relationship between these two variables, we are departing from the classical assumption that real and nominal variables can be studied separately.

We analyze fluctuations in the economy as a whole with the **model of aggregate demand and aggregate supply**, which is illustrated in Figure 2. On the vertical axis is the overall price level in the economy. On the horizontal axis is the overall quantity of goods and services produced in the economy. The **aggregate-demand curve** shows the quantity of goods and services that households, firms, the government, and customers abroad want to buy at each price level. The **aggregate-supply curve** shows the quantity of goods and services that firms produce and sell at each price level. According to this model, the price level and the quantity of output adjust to bring aggregate demand and aggregate supply into balance.

It is tempting to view the model of aggregate demand and aggregate supply as nothing more than a large version of the model of market demand and market supply introduced in Chapter 4. In fact, this model is quite different. When we consider demand and supply in a specific market—ice cream, for instance—the behavior of buyers and sellers depends on the ability of resources to move from one market to another. When the price of ice cream rises, the quantity demanded falls because buyers will use their incomes to buy products other than ice cream. Similarly, a higher price of ice cream raises the quantity supplied because firms that produce ice cream can increase production by hiring workers away from other parts of the economy. This *microeconomic* substitution from one market to another...

---

**Figure 2**

**Aggregate Demand and Aggregate Supply**

Economists use the model of aggregate demand and aggregate supply to analyze economic fluctuations. On the vertical axis is the overall level of prices. On the horizontal axis is the economy’s total output of goods and services. Output and the price level adjust to the point at which the aggregate-supply and aggregate-demand curves intersect.
is impossible for the economy as a whole. After all, the quantity that our model is trying to explain—real GDP—measures the total quantity of goods and services produced by all firms in all markets. To understand why the aggregate-demand curve is downward sloping and why the aggregate-supply curve is upward sloping, we need a macroeconomic theory that explains the total quantity of goods and services demanded and the total quantity of goods and services supplied. Developing such a theory is our next task.

**QUICK QUIZ** How does the economy’s behavior in the short run differ from its behavior in the long run? • Draw the model of aggregate demand and aggregate supply. What variables are on the two axes?

## THE AGGREGATE-DEMAND CURVE

The aggregate-demand curve tells us the quantity of all goods and services demanded in the economy at any given price level. As Figure 3 illustrates, the aggregate-demand curve is downward sloping. This means that, other things equal, a decrease in the economy’s overall level of prices (from, say, \( P_1 \) to \( P_2 \)) raises the quantity of goods and services demanded (from \( Y_1 \) to \( Y_2 \)). Conversely, an increase in the price level reduces the quantity of goods and services demanded.

### WHY THE AGGREGATE-DEMAND CURVE SLOPES DOWNWARD

Why does a change in the price level move the quantity of goods and services demanded in the opposite direction? To answer this question, it is useful to recall

---

**FIGURE 3**

The Aggregate-Demand Curve

A fall in the price level from \( P_1 \) to \( P_2 \) increases the quantity of goods and services demanded from \( Y_1 \) to \( Y_2 \). There are three reasons for this negative relationship. As the price level falls, real wealth rises, interest rates fall, and the exchange rate depreciates. These effects stimulate spending on consumption, investment, and net exports. Increased spending on any or all of these components of output means a larger quantity of goods and services demanded.
that an economy’s GDP (which we denote as $Y$) is the sum of its consumption ($C$), investment ($I$), government purchases ($G$), and net exports ($NX$):

$$Y = C + I + G + NX.$$ 

Each of these four components contributes to the aggregate demand for goods and services. For now, we assume that government spending is fixed by policy. The other three components of spending—consumption, investment, and net exports—depend on economic conditions and, in particular, on the price level. Therefore, to understand the downward slope of the aggregate-demand curve, we must examine how the price level affects the quantity of goods and services demanded for consumption, investment, and net exports.

**The Price Level and Consumption: The Wealth Effect**

Consider the money that you hold in your wallet and your bank account. The nominal value of this money is fixed: One dollar is always worth one dollar. Yet the real value of a dollar is not fixed. If a candy bar costs 1 dollar, then a dollar is worth one candy bar. If the price of a candy bar falls to 50 cents, then 1 dollar is worth two candy bars. Thus, when the price level falls, the dollars you are holding rise in value, which increases your real wealth and your ability to buy goods and services.

This logic gives us the first reason the aggregate demand curve is downward sloping. A decrease in the price level raises the real value of money and makes consumers wealthier, which in turn encourages them to spend more. The increase in consumer spending means a larger quantity of goods and services demanded. Conversely, an increase in the price level reduces the real value of money and makes consumers poorer, which in turn reduces consumer spending and the quantity of goods and services demanded.

**The Price Level and Investment: The Interest-Rate Effect**

The price level is one determinant of the quantity of money demanded. When the price level is lower, households need to hold less money to buy the goods and services they want. Therefore, when the price level falls, households try to reduce their holdings of money by lending some of it out. For instance, a household might use its excess money to buy interest-bearing bonds. Or it might deposit its excess money in an interest-bearing savings account, and the bank would use these funds to make more loans. In either case, as households try to convert some of their money into interest-bearing assets, they drive down interest rates. (The next chapter analyzes this process in more detail.)

Interest rates, in turn, affect spending on goods and services. Because a lower interest rate makes borrowing less expensive, it encourages firms to borrow more to invest in new plants and equipment, and it encourages households to borrow more to invest in new housing. (A lower interest rate might also stimulate consumer spending, especially spending on large durable purchases such as cars, which are often bought on credit.) Thus, a lower interest rate increases the quantity of goods and services demanded.

This logic gives us a second reason the aggregate demand curve is downward sloping. A lower price level reduces the interest rate, encourages greater spending on investment goods, and thereby increases the quantity of goods and services demanded. Conversely, a higher price level raises the interest rate, discourages investment spending, and decreases the quantity of goods and services demanded.

**The Price Level and Net Exports: The Exchange-Rate Effect**

As we have just discussed, a lower price level in the United States lowers the U.S. interest rate. In response to the lower interest rate, some U.S. investors will seek higher returns by
investing abroad. For instance, as the interest rate on U.S. government bonds falls, a mutual fund might sell U.S. government bonds to buy German government bonds. As the mutual fund tries to convert its dollars into euros to buy the German bonds, it increases the supply of dollars in the market for foreign-currency exchange.

The increased supply of dollars to be turned into euros causes the dollar to depreciate relative to the euro. This leads to a change in the real exchange rate—the relative price of domestic and foreign goods. Because each dollar buys fewer units of foreign currencies, foreign goods become more expensive relative to domestic goods.

The change in relative prices affects spending, both at home and abroad. Because foreign goods are now more expensive, Americans buy less from other countries, causing U.S. imports of goods and services to decrease. At the same time, because U.S. goods are now cheaper, foreigners buy more from the United States, so U.S. exports increase. Net exports equal exports minus imports, so both of these changes cause U.S. net exports to increase. Thus, the fall in the real exchange value of the dollar leads to an increase in the quantity of goods and services demanded.

This logic yields a third reason the aggregate demand curve is downward sloping. When a fall in the U.S. price level causes U.S. interest rates to fall, the real value of the dollar declines in foreign exchange markets. This depreciation stimulates U.S. net exports and thereby increases the quantity of goods and services demanded. Conversely, when the U.S. price level rises and causes U.S. interest rates to rise, the real value of the dollar increases, and this appreciation reduces U.S. net exports and the quantity of goods and services demanded.

**Summing Up** There are three distinct but related reasons a fall in the price level increases the quantity of goods and services demanded:

1. Consumers are wealthier, which stimulates the demand for consumption goods.
2. Interest rates fall, which stimulates the demand for investment goods.
3. The currency depreciates, which stimulates the demand for net exports.

The same three effects work in reverse: When the price level rises, decreased wealth depresses consumer spending, higher interest rates depress investment spending, and a currency appreciation depresses net exports.

Here is a thought experiment to hone your intuition about these effects. Imagine that one day you wake up and notice that, for some mysterious reason, the prices of all goods and services have fallen by half, so the dollars you are holding are worth twice as much. In real terms, you now have twice as much money as you had when you went to bed the night before. What would you do with the extra money? You could spend it at your favorite restaurant, increasing consumer spending. You could lend it out (by buying a bond or depositing it in your bank), reducing interest rates and increasing investment spending. Or you could invest it overseas (by buying shares in an international mutual fund), reducing the real exchange value of the dollar and increasing net exports. Whichever of these three responses you choose, the fall in the price level leads to an increase in the quantity of goods and services demanded. This is what the downward slope of the aggregate-demand curve represents.

It is important to keep in mind that the aggregate-demand curve (like all demand curves) is drawn holding “other things equal.” In particular, our three
explanations of the downward-sloping aggregate-demand curve assume that
the money supply is fixed. That is, we have been considering how a change in
the price level affects the demand for goods and services, holding the amount
of money in the economy constant. As we will see, a change in the quantity of
money shifts the aggregate-demand curve. At this point, just keep in mind that
the aggregate-demand curve is drawn for a given quantity of the money supply.

**Why the Aggregate-Demand Curve Might Shift**

The downward slope of the aggregate-demand curve shows that a fall in the price
level raises the overall quantity of goods and services demanded. Many other fac-
tors, however, affect the quantity of goods and services demanded at a given price
level. When one of these other factors changes, the quantity of goods and services
demanded at every price level changes, and the aggregate-demand curve shifts.

Let’s consider some examples of events that shift aggregate demand. We can
categorize them according to which component of spending is most directly
affected.

**Shifts Arising from Changes in Consumption** Suppose Americans suddenly
become more concerned about saving for retirement and, as a result, reduce their
current consumption. Because the quantity of goods and services demanded at
any price level is lower, the aggregate-demand curve shifts to the left. Conversely,
image that a stock-market boom makes people wealthier and less concerned
about saving. The resulting increase in consumer spending means a greater quan-
tity of goods and services demanded at any given price level, so the aggregate-
demand curve shifts to the right.

Thus, any event that changes how much people want to consume at a given
price level shifts the aggregate-demand curve. One policy variable that has this
effect is the level of taxation. When the government cuts taxes, it encourages peo-
ple to spend more, so the aggregate-demand curve shifts to the right. When the
government raises taxes, people cut back on their spending, and the aggregate-
demand curve shifts to the left.

**Shifts Arising from Changes in Investment** Any event that changes how
much firms want to invest at a given price level also shifts the aggregate-demand
curve. For instance, imagine that the computer industry introduces a faster line of
computers, and many firms decide to invest in new computer systems. Because
the quantity of goods and services demanded at any price level is higher, the
aggregate-demand curve shifts to the right. Conversely, if firms become pessimis-
tic about future business conditions, they may cut back on investment spending,
shifting the aggregate-demand curve to the left.

Tax policy can also influence aggregate demand through investment. As we
saw in Chapter 18, an investment tax credit (a tax rebate tied to a firm’s invest-
ment spending) increases the quantity of investment goods that firms demand
at any given interest rate. It therefore shifts the aggregate-demand curve to the
right. The repeal of an investment tax credit reduces investment and shifts the
aggregate-demand curve to the left.

Another policy variable that can influence investment and aggregate demand is
the money supply. As we discuss more fully in the next chapter, an increase in the
money supply lowers the interest rate in the short run. This decrease in the inter-
est rate makes borrowing less costly, which stimulates investment spending and
thereby shifts the aggregate-demand curve to the right. Conversely, a decrease in the money supply raises the interest rate, discourages investment spending, and thereby shifts the aggregate-demand curve to the left. Many economists believe that throughout U.S. history, changes in monetary policy have been an important source of shifts in aggregate demand.

**The 2008 Fiscal Stimulus**

In 2008, the U.S. economy was experiencing a slowdown many economists feared would turn into a recession. To bolster aggregate demand, Congress passed a tax rebate.

**Tax Rebates in $168 Billion Stimulus Plan Begin Arriving in Bank Accounts**

*By John Sullivan*

The federal government began issuing electronic tax rebates on Monday under a $168 billion program to bolster the economy.

President Bush and Congress hope the plan will generate spending and jump-start the economy by putting money in people’s hands. The plan, which was passed by Congress in February, also includes tax breaks for businesses.

Economists have debated how much of an impact the stimulus plan will have. A study of similar rebates in 2001, performed by Joel Slemrod and Matthew D. Shapiro, economics professors at the University of Michigan, found that taxpayers primarily saved the money or used it to pay off debts.

The Treasury Department plans to send electronic rebates to nearly 7.7 million people by the end of this week, said Andrew DeSouza, a spokesman for the department. The government intends to begin mailing checks on May 9 and expects to send a total of about 130 million rebates . . . .

The plan provides rebates of up to $600 for individuals and up to $1,200 for couples filing jointly, with additional payments to families of $300 per child . . . .

For Phylishia White, the rebate check will be a measure of help at a difficult time. Ms. White, 26, of Los Angeles, said her family rented part of a duplex that is now in foreclosure. Facing eviction, they have to move on short notice, and Ms. White said the rebate would help with some of the expenses.

“It’s going to go to something important,” said Ms. White, an administrative assistant. “To the move, to clothes, to food. Probably to the move mostly.”

Professor Slemrod, the Michigan researcher, said Monday that because the tax rebate plan was intended to increase spending rapidly, the tendency among consumers to save or reduce debt with such payments would lessen its impact.

He said it mattered little to the economy what type of purchases—whether groceries or a new washing machine—were made with money from the rebates. What is important, he said, is that the money does go into the economy.

**Shifts Arising from Changes in Government Purchases**  The most direct way that policymakers shift the aggregate-demand curve is through government purchases. For example, suppose Congress decides to reduce purchases of new weapons systems. Because the quantity of goods and services demanded at any price level is lower, the aggregate-demand curve shifts to the left. Conversely, if state governments start building more highways, the result is a greater quantity of goods and services demanded at any price level, so the aggregate-demand curve shifts to the right.

**Shifts Arising from Changes in Net Exports**  Any event that changes net exports for a given price level also shifts aggregate demand. For instance, when Europe experiences a recession, it buys fewer goods from the United States. This reduces U.S. net exports at every price level and shifts the aggregate-demand curve for the U.S. economy to the left. When Europe recovers from its recession, it starts buying U.S. goods again, and the aggregate-demand curve shifts to the right.

Net exports sometimes change because international speculators cause movements in the exchange rate. Suppose, for instance, that these speculators lose confidence in foreign economies and want to move some of their wealth into the U.S. economy. In doing so, they bid up the value of the U.S. dollar in the foreign exchange market. This appreciation of the dollar makes U.S. goods more expensive compared to foreign goods, which depresses net exports and shifts the aggregate-demand curve to the left. Conversely, speculation that causes a depreciation of the dollar stimulates net exports and shifts the aggregate-demand curve to the right.

**Summing Up**  In the next chapter, we analyze the aggregate-demand curve in more detail. There we examine more precisely how the tools of monetary and fiscal policy can shift aggregate demand and whether policymakers should use these tools for that purpose. At this point, however, you should have some idea about why the aggregate-demand curve slopes downward and what kinds of events and policies can shift this curve. Table 1 summarizes what we have learned so far.

**Quick Quiz**  Explain the three reasons the aggregate-demand curve slopes downward.

- Give an example of an event that would shift the aggregate-demand curve. Which way would this event shift the curve?

---

**THE AGGREGATE-SUPPLY CURVE**

The aggregate-supply curve tells us the total quantity of goods and services that firms produce and sell at any given price level. Unlike the aggregate-demand curve, which is always downward sloping, the aggregate-supply curve shows a relationship that depends crucially on the time horizon examined. In the long run, the aggregate-supply curve is vertical, whereas in the short run, the aggregate-supply curve is upward sloping. To understand short-run economic fluctuations, and how the short-run behavior of the economy deviates from its long-run behavior, we need to examine both the long-run aggregate-supply curve and the short-run aggregate-supply curve.
WHY THE AGGREGATE-SUPPLY CURVE IS VERTICAL IN THE LONG RUN

What determines the quantity of goods and services supplied in the long run? We implicitly answered this question earlier in the book when we analyzed the process of economic growth. In the long run, an economy’s production of goods and services (its real GDP) depends on its supplies of labor, capital, and natural resources and on the available technology used to turn these factors of production into goods and services.

When we analyzed these forces that govern long-run growth, we did not need to make any reference to the overall level of prices. We examined the price level in a separate chapter, where we saw that it was determined by the quantity of money. We learned that if two economies were identical except that one had twice as much money in circulation as the other, the price level would be twice as high in the economy with more money. But since the amount of money does not affect technology or the supplies of labor, capital, and natural resources, the output of goods and services in the two economies would be the same.
Because the price level does not affect the long-run determinants of real GDP, the long-run aggregate-supply curve is vertical, as in Figure 4. In other words, in the long run, the economy’s labor, capital, natural resources, and technology determine the total quantity of goods and services supplied, and this quantity supplied is the same regardless of what the price level happens to be.

The vertical long-run aggregate-supply curve is a graphical representation of the classical dichotomy and monetary neutrality. As we have already discussed, classical macroeconomic theory is based on the assumption that real variables do not depend on nominal variables. The long-run aggregate-supply curve is consistent with this idea because it implies that the quantity of output (a real variable) does not depend on the level of prices (a nominal variable). As noted earlier, most economists believe this principle works well when studying the economy over a period of many years but not when studying year-to-year changes. Thus, the aggregate-supply curve is vertical only in the long run.

**WHY THE LONG-RUN AGGREGATE-SUPPLY CURVE MIGHT SHIFT**

Because classical macroeconomic theory predicts the quantity of goods and services produced by an economy in the long run, it also explains the position of the long-run aggregate-supply curve. The long-run level of production is sometimes called *potential output* or *full-employment output*. To be more precise, we call it the **natural rate of output** because it shows what the economy produces when unemployment is at its natural, or normal, rate. The natural rate of output is the level of production toward which the economy gravitates in the long run.

---

natural rate of output
the production of goods and services that an economy achieves in the long run when unemployment is at its normal rate

---

**FIGURE 4**

The Long-Run Aggregate-Supply Curve
In the long run, the quantity of output supplied depends on the economy’s quantities of labor, capital, and natural resources and on the technology for turning these inputs into output. Because the quantity supplied does not depend on the overall price level, the long-run aggregate-supply curve is vertical at the natural rate of output.
Any change in the economy that alters the natural rate of output shifts the long-run aggregate-supply curve. Because output in the classical model depends on labor, capital, natural resources, and technological knowledge, we can categorize shifts in the long-run aggregate-supply curve as arising from these four sources.

**Shifts Arising from Changes in Labor** Imagine that an economy experiences an increase in immigration. Because there would be a greater number of workers, the quantity of goods and services supplied would increase. As a result, the long-run aggregate-supply curve would shift to the right. Conversely, if many workers left the economy to go abroad, the long-run aggregate-supply curve would shift to the left.

The position of the long-run aggregate-supply curve also depends on the natural rate of unemployment, so any change in the natural rate of unemployment shifts the long-run aggregate-supply curve. For example, if Congress were to raise the minimum wage substantially, the natural rate of unemployment would rise, and the economy would produce a smaller quantity of goods and services. As a result, the long-run aggregate-supply curve would shift to the left. Conversely, if a reform of the unemployment insurance system were to encourage unemployed workers to search harder for new jobs, the natural rate of unemployment would fall, and the long-run aggregate-supply curve would shift to the right.

**Shifts Arising from Changes in Capital** An increase in the economy’s capital stock increases productivity and, thereby, the quantity of goods and services supplied. As a result, the long-run aggregate-supply curve shifts to the right. Conversely, a decrease in the economy’s capital stock decreases productivity and the quantity of goods and services supplied, shifting the long-run aggregate-supply curve to the left.

Notice that the same logic applies regardless of whether we are discussing physical capital such as machines and factories or human capital such as college degrees. An increase in either type of capital will raise the economy’s ability to produce goods and services and, thus, shift the long-run aggregate-supply curve to the right.

**Shifts Arising from Changes in Natural Resources** An economy’s production depends on its natural resources, including its land, minerals, and weather. A discovery of a new mineral deposit shifts the long-run aggregate-supply curve to the right. A change in weather patterns that makes farming more difficult shifts the long-run aggregate-supply curve to the left.

In many countries, important natural resources are imported. A change in the availability of these resources can also shift the aggregate-supply curve. As we discuss later in this chapter, events occurring in the world oil market have historically been an important source of shifts in aggregate supply for the United States and other oil-importing nations.

**Shifts Arising from Changes in Technological Knowledge** Perhaps the most important reason that the economy today produces more than it did a generation ago is that our technological knowledge has advanced. The invention of the computer, for instance, has allowed us to produce more goods and services from any given amounts of labor, capital, and natural resources. As computer use has spread throughout the economy, it has shifted the long-run aggregate-supply curve to the right.
Although not literally technological, many other events act like changes in technology. For instance, opening up international trade has effects similar to inventing new production processes because it allows a country to specialize in higher-productivity industries, so it also shifts the long-run aggregate-supply curve to the right. Conversely, if the government passes new regulations preventing firms from using some production methods, perhaps to address worker safety or environmental concerns, the result would be a leftward shift in the long-run aggregate-supply curve.

**Summing Up** Because the long-run aggregate-supply curve reflects the classical model of the economy we developed in previous chapters, it provides a new way to describe our earlier analysis. Any policy or event that raised real GDP in previous chapters can now be described as increasing the quantity of goods and services supplied and shifting the long-run aggregate-supply curve to the right. Any policy or event that lowered real GDP in previous chapters can now be described as decreasing the quantity of goods and services supplied and shifting the long-run aggregate-supply curve to the left.

**Using Aggregate Demand and Aggregate Supply to Depict Long-Run Growth and Inflation**

Having introduced the economy’s aggregate-demand curve and the long-run aggregate-supply curve, we now have a new way to describe the economy’s long-run trends. Figure 5 illustrates the changes that occur in an economy from decade to decade. Notice that both curves are shifting. Although there are many forces that govern the economy in the long run and can in theory cause such shifts, the two most important in the real world are technology and monetary policy. Technological progress enhances an economy’s ability to produce goods and services, and this increase in output is reflected in the continual shifts of the long-run aggregate-supply curve to the right. At the same time, because the Fed increases the money supply over time, the aggregate-demand curve also shifts to the right. As the figure illustrates, the result is continuing growth in output (as shown by increasing \(Y\)) and continuing inflation (as shown by increasing \(P\)). This is just another way of representing the classical analysis of growth and inflation we conducted in earlier chapters.

The purpose of developing the model of aggregate demand and aggregate supply, however, is not to dress our previous long-run conclusions in new clothing. Instead, it is to provide a framework for short-run analysis, as we will see in a moment. As we develop the short-run model, we keep the analysis simple by not showing the continuing growth and inflation depicted by the shifts in Figure 5. But always remember that long-run trends are the background upon which short-run fluctuations are superimposed. *Short-run fluctuations in output and the price level should be viewed as deviations from the continuing long-run trends of output growth and inflation.*

**Why the Aggregate-Supply Curve Slopes Upward in the Short Run**

The key difference between the economy in the short run and in the long run is the behavior of aggregate supply. The long-run aggregate-supply curve is vertical
because, in the long run, the overall level of prices does not affect the economy’s ability to produce goods and services. By contrast, in the short run, the price level does affect the economy’s output. That is, over a period of a year or two, an increase in the overall level of prices in the economy tends to raise the quantity of goods and services supplied, and a decrease in the level of prices tends to reduce the quantity of goods and services supplied. As a result, the short-run aggregate-supply curve is upward sloping, as shown in Figure 6.

Why do changes in the price level affect output in the short run? Macro-economists have proposed three theories for the upward slope of the short-run aggregate-supply curve. In each theory, a specific market imperfection causes the supply side of the economy to behave differently in the short run than it does in
the long run. The following theories differ in their details, but they share a common theme: The quantity of output supplied deviates from its long-run, or natural, level when the actual price level in the economy deviates from the price level that people expected to prevail. When the price level rises above the level that people expected, output rises above its natural rate, and when the price level falls below the expected level, output falls below its natural rate.

**The Sticky-Wage Theory** The first explanation of the upward slope of the short-run aggregate-supply curve is the sticky-wage theory. Because this theory is the simplest of the three approaches to aggregate supply, it is the one we emphasize in this book.

According to this theory, the short-run aggregate-supply curve slopes upward because nominal wages are slow to adjust to changing economic conditions. In other words, wages are “sticky” in the short run. To some extent, the slow adjustment of nominal wages is attributable to long-term contracts between workers and firms that fix nominal wages, sometimes for as long as three years. In addition, this prolonged adjustment may be attributable to slowly changing social norms and notions of fairness that influence wage setting.

An example helps explain how sticky nominal wages can result in a short-run aggregate-supply curve that slopes upward. Imagine that a year ago a firm expected the price level to be 100, and based on this expectation, it signed a contract with its workers agreeing to pay them, say, $20 an hour. In fact, the price level, $P$, turns out to be only 95. Because prices have fallen below expectations, the firm gets 5 percent less than expected for each unit of its product that it sells. The cost of labor used to make the output, however, is stuck at $20 per hour. Production is now less profitable, so the firm hires fewer workers and reduces the quantity of output supplied. Over time, the labor contract will expire, and the firm can renegotiate with its workers for a lower wage (which they may accept because...
prices are lower), but in the meantime, employment and production will remain below their long-run levels.

The same logic works in reverse. Suppose the price level turns out to be 105, and the wage remains stuck at $20. The firm sees that the amount it is paid for each unit sold is up by 5 percent, while its labor costs are not. In response, it hires more workers and increases the quantity supplied. Eventually, the workers will demand higher nominal wages to compensate for the higher price level, but for a while, the firm can take advantage of the profit opportunity by increasing employment and the quantity of output supplied above their long-run levels.

In short, according to the sticky-wage theory, the short-run aggregate-supply curve is upward sloping because nominal wages are based on expected prices and do not respond immediately when the actual price level turns out to be different from what was expected. This stickiness of wages gives firms an incentive to produce less output when the price level turns out lower than expected and to produce more when the price level turns out higher than expected.

The **Sticky-Price Theory** Some economists have advocated another approach to explaining the upward slope of the short-run aggregate-supply curve, called the sticky-price theory. As we just discussed, the sticky-wage theory emphasizes that nominal wages adjust slowly over time. The sticky-price theory emphasizes that the prices of some goods and services also adjust sluggishly in response to changing economic conditions. This slow adjustment of prices occurs in part because there are costs to adjusting prices, called *menu costs*. These menu costs include the cost of printing and distributing catalogs and the time required to change price tags. As a result of these costs, prices as well as wages may be sticky in the short run.

To see how sticky prices explain the aggregate-supply curve’s upward slope, suppose that each firm in the economy announces its prices in advance based on the economic conditions it expects to prevail over the coming year. Suppose further that after prices are announced, the economy experiences an unexpected contraction in the money supply, which (as we have learned) will reduce the overall price level in the long run. Although some firms can reduce their prices immediately in response to an unexpected change in economic conditions, other firms may not want to incur additional menu costs. As a result, they may temporarily lag behind in reducing their prices. Because these lagging firms have prices that are too high, their sales decline. Declining sales, in turn, cause these firms to cut back on production and employment. In other words, because not all prices adjust instantly to changing economic conditions, an unexpected fall in the price level leaves some firms with higher-than-desired prices, and these higher-than-desired prices depress sales and induce firms to reduce the quantity of goods and services they produce.

The same reasoning applies when the money supply and price level turn out to be above what firms expected when they originally set their prices. While some firms raise their prices immediately in response to the new economic environment, other firms lag behind, keeping their prices at the lower-than-desired levels. These low prices attract customers, which induces these firms to increase employment and production. Thus, during the time these lagging firms are operating with outdated prices, there is a positive association between the overall price level and the quantity of output. This positive association is represented by the upward slope of the short-run aggregate-supply curve.
The Misperceptions Theory  A third approach to explaining the upward slope of the short-run aggregate-supply curve is the misperceptions theory. According to this theory, changes in the overall price level can temporarily mislead suppliers about what is happening in the individual markets in which they sell their output. As a result of these short-run misperceptions, suppliers respond to changes in the level of prices, and this response leads to an upward-sloping aggregate-supply curve.

To see how this might work, suppose the overall price level falls below the level that suppliers expected. When suppliers see the prices of their products fall, they may mistakenly believe that their relative prices have fallen; that is, they may believe that their prices have fallen compared to other prices in the economy. For example, wheat farmers may notice a fall in the price of wheat before they notice a fall in the prices of the many items they buy as consumers. They may infer from this observation that the reward to producing wheat is temporarily low, and they may respond by reducing the quantity of wheat they supply. Similarly, workers may notice a fall in their nominal wages before they notice that the prices of the goods they buy are also falling. They may infer that the reward for working is temporarily low and respond by reducing the quantity of labor they supply. In both cases, a lower price level causes misperceptions about relative prices, and these misperceptions induce suppliers to respond to the lower price level by decreasing the quantity of goods and services supplied.

Similar misperceptions arise when the price level is above what was expected. Suppliers of goods and services may notice the price of their output rising and infer, mistakenly, that their relative prices are rising. They would conclude that it is a good time to produce. Until their misperceptions are corrected, they respond to the higher price level by increasing the quantity of goods and services supplied. This behavior results in a short-run aggregate-supply curve that slopes upward.

Summing Up  There are three alternative explanations for the upward slope of the short-run aggregate-supply curve: (1) sticky wages, (2) sticky prices, and (3) misperceptions about relative prices. Economists debate which of these theories is correct, and it is very possible each contains an element of truth. For our purposes in this book, the similarities of the theories are more important than the differences. All three theories suggest that output deviates in the short run from its long-run level (the natural rate) when the actual price level deviates from the price level that people had expected to prevail. We can express this mathematically as follows:

\[
\text{Quantity of output} = \text{Natural rate of output} + a \left( \frac{\text{Actual price level} - \text{Expected price level}}{\text{price level}} \right)
\]

where \( a \) is a number that determines how much output responds to unexpected changes in the price level.

Notice that each of the three theories of short-run aggregate supply emphasizes a problem that is likely to be temporary. Whether the upward slope of the aggregate-supply curve is attributable to sticky wages, sticky prices, or misperceptions, these conditions will not persist forever. Over time, nominal wages will become unstuck, prices will become unstuck, and misperceptions about relative prices will be corrected. In the long run, it is reasonable to assume that wages
and prices are flexible rather than sticky and that people are not confused about relative prices. Thus, while we have several good theories to explain why the short-run aggregate-supply curve is upward sloping, they are all consistent with a long-run aggregate-supply curve that is vertical.

**Why the Short-Run Aggregate-Supply Curve Might Shift**

The short-run aggregate-supply curve tells us the quantity of goods and services supplied in the short run for any given level of prices. This curve is similar to the long-run aggregate-supply curve, but it is upward sloping rather than vertical because of sticky wages, sticky prices, and misperceptions. Thus, when thinking about what shifts the short-run aggregate-supply curve, we have to consider all those variables that shift the long-run aggregate-supply curve plus a new variable—the expected price level—that influences the wages that are stuck, the prices that are stuck, and the perceptions about relative prices that may be flawed.

Let’s start with what we know about the long-run aggregate-supply curve. As we discussed earlier, shifts in the long-run aggregate-supply curve normally arise from changes in labor, capital, natural resources, or technological knowledge. These same variables shift the short-run aggregate-supply curve. For example, when an increase in the economy’s capital stock increases productivity, the economy is able to produce more output, so both the long-run and short-run aggregate-supply curves shift to the right. When an increase in the minimum wage raises the natural rate of unemployment, the economy has fewer employed workers and thus produces less output, so both the long-run and short-run aggregate-supply curves shift to the left.

The important new variable that affects the position of the short-run aggregate-supply curve is the price level that people expected to prevail. As we have discussed, the quantity of goods and services supplied depends, in the short run, on sticky wages, sticky prices, and misperceptions. Yet wages, prices, and perceptions are set based on the expected price level. So when people change their expectations of the price level, the short-run aggregate-supply curve shifts.

To make this idea more concrete, let’s consider a specific theory of aggregate supply—the sticky-wage theory. According to this theory, when workers and firms expect the price level to be high, they are more likely to reach a bargain with a high level of nominal wages. High wages raise firms’ costs, and for any given actual price level, higher costs reduce the quantity of goods and services supplied. Thus, when the expected price level rises, wages are higher, costs increase, and firms produce a smaller quantity of goods and services at any given actual price level. Thus, the short-run aggregate-supply curve shifts to the left. Conversely, when the expected price level falls, wages are lower, costs decline, firms increase output at any given price level, and the short-run aggregate-supply curve shifts to the right.

A similar logic applies in each theory of aggregate supply. The general lesson is the following: An increase in the expected price level reduces the quantity of goods and services supplied and shifts the short-run aggregate-supply curve to the left. A decrease in the expected price level raises the quantity of goods and services supplied and shifts the short-run aggregate-supply curve to the right. As we will see in the next section, the influence of expectations on the position of the short-run aggregate-supply curve plays a key role in explaining how the economy makes the transition from the short run to the long run. In the short run, expectations are fixed, and the
Now that we have introduced the model of aggregate demand and aggregate supply, we have the basic tools we need to analyze fluctuations in economic activity. In particular, we can use what we have learned about aggregate demand and aggregate supply to examine the two basic causes of short-run fluctuations: shifts in aggregate demand and shifts in aggregate supply.

**Why Does the Short-Run Aggregate-Supply Curve Slope Upward?**

1. **The Sticky-Wage Theory**: An unexpectedly low price level raises the real wage, which causes firms to hire fewer workers and produce a smaller quantity of goods and services.
2. **The Sticky-Price Theory**: An unexpectedly low price level leaves some firms with higher-than-desired prices, which depresses their sales and leads them to cut back production.
3. **The Misperceptions Theory**: An unexpectedly low price level leads some suppliers to think their relative prices have fallen, which induces a fall in production.

**Why Might the Short-Run Aggregate-Supply Curve Shift?**

1. **Shifts Arising from Labor**: An increase in the quantity of labor available (perhaps due to a fall in the natural rate of unemployment) shifts the aggregate-supply curve to the right. A decrease in the quantity of labor available (perhaps due to a rise in the natural rate of unemployment) shifts the aggregate-supply curve to the left.
2. **Shifts Arising from Capital**: An increase in physical or human capital shifts the aggregate-supply curve to the right. A decrease in physical or human capital shifts the aggregate-supply curve to the left.
3. **Shifts Arising from Natural Resources**: An increase in the availability of natural resources shifts the aggregate-supply curve to the right. A decrease in the availability of natural resources shifts the aggregate-supply curve to the left.
4. **Shifts Arising from Technology**: An advance in technological knowledge shifts the aggregate-supply curve to the right. A decrease in the available technology (perhaps due to government regulation) shifts the aggregate-supply curve to the left.
5. **Shifts Arising from the Expected Price Level**: A decrease in the expected price level shifts the short-run aggregate-supply curve to the right. An increase in the expected price level shifts the short-run aggregate-supply curve to the left.

---

**TABLE 2**

**TWO CAUSES OF ECONOMIC FLUCTUATIONS**

Now that we have introduced the model of aggregate demand and aggregate supply, we have the basic tools we need to analyze fluctuations in economic activity. In particular, we can use what we have learned about aggregate demand and aggregate supply to examine the two basic causes of short-run fluctuations: shifts in aggregate demand and shifts in aggregate supply.

**QUICK QUIZ**

Explain why the long-run aggregate-supply curve is vertical. • Explain three theories for why the short-run aggregate-supply curve is upward sloping. • What variables shift both the long-run and short-run aggregate-supply curves? • What variable shifts the short-run aggregate-supply curve but not the long-run aggregate-supply curve?
To keep things simple, we assume the economy begins in long-run equilibrium, as shown in Figure 7. Output and the price level are determined in the long run by the intersection of the aggregate-demand curve and the long-run aggregate-supply curve, shown as point A in the figure. At this point, output is at its natural rate. Because the economy is always in a short-run equilibrium, the short-run aggregate-supply curve passes through this point as well, indicating that the expected price level has adjusted to this long-run equilibrium. That is, when an economy is in its long-run equilibrium, the expected price level must equal the actual price level so that the intersection of aggregate demand with short-run aggregate supply is the same as the intersection of aggregate demand with long-run aggregate supply.

**The Effects of a Shift in Aggregate Demand**

Suppose that a wave of pessimism suddenly overtakes the economy. The cause might be a scandal in the White House, a crash in the stock market, or the outbreak of war overseas. Because of this event, many people lose confidence in the future and alter their plans. Households cut back on their spending and delay major purchases, and firms put off buying new equipment.

What is the macroeconomic impact of such a wave of pessimism? In answering this question, we can follow the three steps we used in Chapter 4 when analyzing supply and demand in specific markets. First, we determine whether the event affects aggregate demand or aggregate supply. Second, we decide which direction the curve shifts. Third, we use the diagram of aggregate demand and aggregate supply to compare the initial and the new equilibrium. The new wrinkle is that we need to add a fourth step: We have to keep track of a new short-run equilibrium, a new long-run equilibrium, and the transition between them. Table 3 summarizes the four steps to analyzing economic fluctuations.
The first two steps are easy. First, because the wave of pessimism affects spending plans, it affects the aggregate-demand curve. Second, because households and firms now want to buy a smaller quantity of goods and services for any given price level, the event reduces aggregate demand. As Figure 8 shows, the aggregate-demand curve shifts to the left from $AD_1$ to $AD_2$.

With this figure, we can perform step three: By comparing the initial and the new equilibrium, we can see the effects of the fall in aggregate demand. In the short run, the economy moves along the initial short-run aggregate-supply curve, $AS_1$, going from point A to point B. As the economy moves between these two points, output falls from $Y_1$ to $Y_2$, and the price level falls from $P_1$ to $P_2$. The falling level of output indicates that the economy is in a recession. Although not shown in the figure, firms respond to lower sales and production by reducing...
employment. Thus, the pessimism that caused the shift in aggregate demand is, to some extent, self-fulfilling: Pessimism about the future leads to falling incomes and rising unemployment.

Now comes step four—the transition from the short-run equilibrium to the long-run equilibrium. Because of the reduction in aggregate demand, the price level initially falls from $P_1$ to $P_2$. The price level is thus below the level that people were expecting ($P_1$) before the sudden fall in aggregate demand. People can be surprised in the short run, but they will not remain surprised. Over time, expectations catch up with this new reality, and the expected price level falls as well. The fall in the expected price level alters wages, prices, and perceptions, which in turn influences the position of the short-run aggregate-supply curve. For example, according to the sticky-wage theory, once workers and firms come to expect a lower level of prices, they start to strike bargains for lower nominal wages; the reduction in labor costs encourages firms to hire more workers and expands production at any given level of prices. Thus, the fall in the expected price level shifts the short-run aggregate-supply curve to the right from $AS_1$ to $AS_2$ in Figure 8. This shift allows the economy to approach point C, where the new aggregate-demand curve ($AD_2$) crosses the long-run aggregate-supply curve.

In the new long-run equilibrium, point C, output is back to its natural rate. The economy has corrected itself: The decline in output is reversed in the long run, even without action by policymakers. Although the wave of pessimism has reduced aggregate demand, the price level has fallen sufficiently (to $P_3$) to offset the shift in the aggregate-demand curve, and people have come to expect this new

---

**FYI**

**Monetary Neutrality Revisited**

According to classical economic theory, money is neutral. That is, changes in the quantity of money affect nominal variables such as the price level but not real variables such as output. Earlier in this chapter, we noted that most economists accept this conclusion as a description of how the economy works in the long run but not in the short run. With the model of aggregate demand and aggregate supply, we can illustrate this conclusion and explain it more fully.

Suppose that the Federal Reserve reduces the quantity of money in the economy. What effect does this change have? As we discussed, the money supply is one determinant of aggregate demand. The reduction in the money supply shifts the aggregate-demand curve to the left.

The analysis looks just like Figure 8. Even though the cause of the shift in aggregate demand is different, we would observe the same effects on output and the price level. In the short run, both output and the price level fall. The economy experiences a recession. But over time, the expected price level falls as well. Firms and workers respond to their new expectations by, for instance, agreeing to lower nominal wages. As they do so, the short-run aggregate-supply curve shifts to the right. Eventually, the economy finds itself back on the long-run aggregate-supply curve.

Figure 8 shows when money matters for real variables and when it does not: In the long run, money is neutral, as represented by the movement of the economy from point A to point C. But in the short run, a change in the money supply has real effects, as represented by the movement of the economy from point A to point B. An old saying summarizes the analysis: “Money is a veil, but when the veil flutters, real output sputters.”
lower price level as well. Thus, in the long run, the shift in aggregate demand is reflected fully in the price level and not at all in the level of output. In other words, the long-run effect of a shift in aggregate demand is a nominal change (the price level is lower) but not a real change (output is the same).

What should policymakers do when faced with a sudden fall in aggregate demand? In this analysis, we assumed they did nothing. Another possibility is that, as soon as the economy heads into recession (moving from point A to point B), policymakers could take action to increase aggregate demand. As we noted earlier, an increase in government spending or an increase in the money supply would increase the quantity of goods and services demanded at any price and, therefore, would shift the aggregate-demand curve to the right. If policymakers act with sufficient speed and precision, they can offset the initial shift in aggregate demand, return the aggregate-demand curve to $AD_1$, and bring the economy back to point A. If the policy is successful, the painful period of depressed output and employment can be reduced in length and severity. The next chapter discusses in more detail the ways in which monetary and fiscal policy influence aggregate demand, as well as some of the practical difficulties in using these policy instruments.

To sum up, this story about shifts in aggregate demand has three important lessons:

- In the short run, shifts in aggregate demand cause fluctuations in the economy’s output of goods and services.
- In the long run, shifts in aggregate demand affect the overall price level but do not affect output.
- Policymakers who influence aggregate demand can potentially mitigate the severity of economic fluctuations.

### Case Study: Two Big Shifts in Aggregate Demand: The Great Depression and World War II

At the beginning of this chapter, we established three key facts about economic fluctuations by looking at data since 1965. Let’s now take a longer look at U.S. economic history. Figure 9 shows data since 1900 on the percentage change in real GDP over the previous 3 years. In an average 3-year period, real GDP grows about 10 percent—a bit more than 3 percent per year. The business cycle, however, causes fluctuations around this average. Two episodes jump out as being particularly significant: the large drop in real GDP in the early 1930s and the large increase in real GDP in the early 1940s. Both of these events are attributable to shifts in aggregate demand.

The economic calamity of the early 1930s is called the Great Depression, and it is by far the largest economic downturn in U.S. history. Real GDP fell by 27 percent from 1929 to 1933, and unemployment rose from 3 percent to 25 percent. At the same time, the price level fell by 22 percent over these 4 years. Many other countries experienced similar declines in output and prices during this period.

Economic historians continue to debate the causes of the Great Depression, but most explanations center on a large decline in aggregate demand. What caused aggregate demand to contract? Here is where the disagreement arises.

Many economists place primary blame on the decline in the money supply: From 1929 to 1933, the money supply fell by 28 percent. As you may recall from
our discussion of the monetary system, this decline in the money supply was due to problems in the banking system. As households withdrew their money from financially shaky banks and bankers became more cautious and started holding greater reserves, the process of money creation under fractional-reserve banking went into reverse. The Fed, meanwhile, failed to offset this fall in the money multiplier with expansionary open-market operations. As a result, the money supply declined. Many economists blame the Fed’s failure to act for the Great Depression’s severity.

Other economists have suggested alternative reasons for the collapse in aggregate demand. For example, stock prices fell about 90 percent during this period, depressing household wealth and thereby consumer spending. In addition, the banking problems may have prevented some firms from obtaining the financing they wanted for investment projects, and this would have depressed investment spending. It is possible that all these forces may have acted together to contract aggregate demand during the Great Depression.

The second significant episode in Figure 9—the economic boom of the early 1940s—is easier to explain. The obvious cause of this event was World War II. As the United States entered the war overseas, the federal government had to devote more resources to the military. Government purchases of goods and services increased almost fivefold from 1939 to 1944. This huge expansion in aggregate demand almost doubled the economy’s production of goods and services and led to a 20 percent increase in the price level (although widespread government price controls limited the rise in prices). Unemployment fell from 17 percent in 1939 to about 1 percent in 1944—the lowest level in U.S. history.
THE RECESSION OF 2001

After the longest economic expansion in history, the U.S. economy experienced a recession in 2001. The unemployment rate rose from 3.9 percent in December 2000 to 4.9 percent in August 2001 and to 6.3 percent in June 2003. The unemployment rate then began to decline. By January 2005, unemployment had fallen back to 5.2 percent.

What caused the recession, and what ended it? The answer to both questions is shifts in aggregate demand.

The recession began with the end of the dot-com bubble in the stock market. During the 1990s, many stock-market investors became optimistic about information technology, and they bid up stock prices, particularly of high-tech companies. With hindsight, it is fair to say that this optimism was excessive. Eventually, the optimism faded, and stock prices fell by about 25 percent from August 2000 to August 2001. The fall in the stock market reduced household wealth, which in turn reduced consumer spending. In addition, when the new technologies started to appear less profitable than they had originally seemed, investment spending fell. The aggregate-demand curve shifted to the left.

The second shock to the economy was the terrorist attacks on New York and Washington on September 11, 2001. In the week after the attacks, the stock market fell another 12 percent, its biggest weekly loss since the Great Depression of the 1930s. Moreover, the attacks increased uncertainty about what the future would hold. Uncertainty can reduce spending, as households and firms postpone plans, waiting for the uncertainty to be resolved. Thus, the terrorist attacks shifted the aggregate-demand curve farther to the left.

The third event that put downward pressure on aggregate demand was a series of corporate accounting scandals. During 2001 and 2002, several major corporations, including Enron and WorldCom, were found to have misled the public about their profitability. When the truth became known, the value of their stock plummeted. Even honest companies experienced stock declines, as stock-market investors became less trustful of all accounting data. This fall in the stock market further depressed aggregate demand.

Policymakers were quick to respond to these events. As soon as the economic slowdown became apparent, the Federal Reserve pursued expansionary monetary policy. Money growth accelerated, and interest rates fell. The federal funds rate (the interest rate on loans between banks that the Fed uses as its short-term policy target) fell from 6.5 percent in December 2000 to 1.0 percent in June 2003. Lower interest rates stimulated spending by reducing the cost of borrowing. At the same time, with the president’s urging, Congress passed a tax cut in 2001, including an immediate tax rebate, and another tax cut in 2003. One goal of these tax cuts was to stimulate consumer and investment spending. Interest-rate cuts and tax cuts both shifted the aggregate-demand curve to the right, offsetting the three contractionary shocks the economy had experienced.

The recession of 2001 is a reminder of the many kinds of events that can influence aggregate demand and, thus, the direction of the economy.

THE EFFECTS OF A SHIFT IN AGGREGATE SUPPLY

Imagine once again an economy in its long-run equilibrium. Now suppose that suddenly some firms experience an increase in their costs of production. For
example, bad weather in farm states might destroy some crops, driving up the cost of producing food products. Or a war in the Middle East might interrupt the shipping of crude oil, driving up the cost of producing oil products.

To analyze the macroeconomic impact of such an increase in production costs, we follow the same four steps. First, which curve is affected? Because production costs affect the firms that supply goods and services, changes in production costs alter the position of the aggregate-supply curve. Second, which direction does the curve shift? Because higher production costs make selling goods and services less profitable, firms now supply a smaller quantity of output for any given price level. Thus, as Figure 10 shows, the short-run aggregate-supply curve shifts to the left from $AS_1$ to $AS_2$. (Depending on the event, the long-run aggregate-supply curve might also shift. To keep things simple, however, we will assume that it does not.)

The figure allows us to perform step three of comparing the initial and the new equilibrium. In the short run, the economy goes from point A to point B, moving along the existing aggregate-demand curve. The output of the economy falls from $Y_1$ to $Y_2$, and the price level rises from $P_1$ to $P_2$. Because the economy is experiencing both stagnation (falling output) and inflation (rising prices), such an event is sometimes called stagflation.

Now consider step four—the transition from the short-run equilibrium to the long-run equilibrium. According to the sticky-wage theory, the key issue is how stagflation affects nominal wages. Firms and workers may at first respond to the higher level of prices by raising their expectations of the price level and setting higher nominal wages. In this case, firms’ costs will rise yet again, and the short-run aggregate-supply curve will shift farther to the left, making the problem of

---

**Figure 10: An Adverse Shift in Aggregate Supply**

When some event increases firms’ costs, the short-run aggregate-supply curve shifts to the left from $AS_1$ to $AS_2$. The economy moves from point A to point B. The result is stagflation: Output falls from $Y_1$ to $Y_2$, and the price level rises from $P_1$ to $P_2$.

1. An adverse shift in the short-run aggregate-supply curve . . .
2. . . . causes output to fall . . .
3. . . . and the price level to rise.

---

**Stagflation**

A period of falling output and rising prices.
stagflation even worse. This phenomenon of higher prices leading to higher wages, in turn leading to even higher prices, is sometimes called a wage-price spiral.

At some point, this spiral of ever-rising wages and prices will slow. The low level of output and employment will put downward pressure on workers’ wages because workers have less bargaining power when unemployment is high. As nominal wages fall, producing goods and services becomes more profitable, and the short-run aggregate-supply curve shifts to the right. As it shifts back toward $AS_1$, the price level falls, and the quantity of output approaches its natural rate. In the long run, the economy returns to point A, where the aggregate-demand curve crosses the long-run aggregate-supply curve.

This transition back to the initial equilibrium assumes, however, that aggregate demand is held constant throughout the process. In the real world, that may not be the case. Policymakers who control monetary and fiscal policy might attempt to offset some of the effects of the shift in the short-run aggregate-supply curve by shifting the aggregate-demand curve. This possibility is shown in Figure 11. In this case, changes in policy shift the aggregate-demand curve to the right from $AD_1$ to $AD_2$—exactly enough to prevent the shift in aggregate supply from affecting output. The economy moves directly from point A to point C. Output remains at its natural rate, and the price level rises from $P_1$ to $P_3$. In this case, policymakers are said to accommodate the shift in aggregate supply. An accommodative policy accepts a permanently higher level of prices to maintain a higher level of output and employment.

To sum up, this story about shifts in aggregate supply has two important lessons:
Shifts in aggregate supply can cause stagflation—a combination of recession (falling output) and inflation (rising prices).

Policymakers who can influence aggregate demand can potentially mitigate the adverse impact on output but only at the cost of exacerbating the problem of inflation.

**OIL AND THE ECONOMY**

Some of the largest economic fluctuations in the U.S. economy since 1970 have originated in the oil fields of the Middle East. Crude oil is a key input into the production of many goods and services, and much of the world’s oil comes from Saudi Arabia, Kuwait, and other Middle Eastern countries. When some event (usually political in origin) reduces the supply of crude oil flowing from this region, the price of oil rises around the world. U.S. firms that produce gasoline, tires, and many other products experience rising costs, and they find it less profitable to supply their output of goods and services at any given price level. The result is a leftward shift in the aggregate-supply curve, which in turn leads to stagflation.
The first episode of this sort occurred in the mid-1970s. The countries with large oil reserves got together as members of OPEC, the Organization of Petroleum Exporting Countries. OPEC is a cartel—a group of sellers that attempts to thwart competition and reduce production to raise prices. And indeed, oil prices rose substantially. From 1973 to 1975, oil approximately doubled in price. Oil-importing countries around the world experienced simultaneous inflation and recession. The U.S. inflation rate as measured by the CPI exceeded 10 percent for the first time in decades. Unemployment rose from 4.9 percent in 1973 to 8.5 percent in 1975.

Almost the same thing happened a few years later. In the late 1970s, the OPEC countries again restricted the supply of oil to raise the price. From 1978 to 1981, the price of oil more than doubled. Once again, the result was stagflation. Inflation, which had subsided somewhat after the first OPEC event, again rose above 10 percent per year. But because the Fed was not willing to accommodate such a large rise in inflation, a recession was soon to follow. Unemployment rose from about 6 percent in 1978 and 1979 to about 10 percent a few years later.

The world market for oil can also be a source of favorable shifts in aggregate supply. In 1986, squabbling broke out among members of OPEC. Member countries reneged on their agreements to restrict oil production. In the world market for crude oil, prices fell by about half. This fall in oil prices reduced costs to U.S. firms, which now found it more profitable to supply goods and services at any given price level. As a result, the aggregate-supply curve shifted to the right. The U.S. economy experienced the opposite of stagflation: Output grew rapidly, unemployment fell, and the inflation rate reached its lowest level in many years.

In recent years, the world market for oil has not been as important a source of economic fluctuations. Part of the reason is that conservation efforts and changes in technology have reduced the economy’s dependence on oil. The amount of oil used to produce a unit of real GDP has declined about 40 percent since the OPEC shocks of the 1970s. As a result, the economic impact of any change in oil prices is smaller today than it was in the past.

Nonetheless, it would be premature to conclude that the United States no longer needs to worry about oil prices. If the rise in oil prices were ever large enough, the macroeconomic result would most likely resemble the events of the 1970s. And indeed in 2008, as this book was going to press, world oil prices were rising significantly, in part because of increased demand from a rapidly growing China. These oil price increases were leading some observers to fear that the U.S. economy may once again suffer from stagflation.

**Quick Quiz** Suppose that the election of a popular presidential candidate suddenly increases people’s confidence in the future. Use the model of aggregate demand and aggregate supply to analyze the effect on the economy.

**Conclusion**

This chapter has achieved two goals. First, we have discussed some of the important facts about short-run fluctuations in economic activity. Second, we have introduced a basic model to explain those fluctuations, called the model of aggregate demand and aggregate supply. We continue our study of this model in the next chapter to understand more fully what causes fluctuations in the economy and how policymakers might respond to these fluctuations.
Three theories have been proposed to explain the upward slope of the short-run aggregate-supply curve. According to the sticky-wage theory, an unexpected fall in the price level temporarily raises real wages, which induces firms to reduce employment and production. According to the sticky-price theory, an unexpected fall in the price level leaves some firms with prices that are temporarily too high, which reduces their sales and causes them to cut back production. According to the misperceptions theory, an unexpected fall in the price level leads suppliers to mistakenly believe that their relative prices have fallen, which induces them to reduce production. All three theories imply that output deviates from its natural rate when the actual price level deviates from the price level that people expected.

Events that alter the economy’s ability to produce output, such as changes in labor, capital, natural resources, or technology, shift the short-run aggregate-supply curve (and may shift the long-run aggregate-supply curve as well). In addition, the position of the short-run aggregate-supply curve depends on the expected price level.

One possible cause of economic fluctuations is a shift in aggregate demand. When the aggregate-demand curve shifts to the left, for instance, output and prices fall in the short run. Over time, as a change in the expected price level causes wages, prices, and perceptions to adjust, the short-run aggregate-supply curve shifts back to the right. This shift returns the economy to its natural rate of output at a new, lower price level.

A second possible cause of economic fluctuations is a shift in aggregate supply. When the short-run aggregate-supply curve shifts to the left, the effect is falling output and rising prices—a combination called stagflation. Over time, as wages, prices, and perceptions adjust, the short-run aggregate-supply curve shifts back to the right, returning the price level and output to their original levels.
KEY CONCEPTS

- recession, p. 507
- depression, p. 507
- model of aggregate demand and aggregate supply, p. 513
- aggregate-demand curve, p. 513
- aggregate-supply curve, p. 513
- natural rate of output, p. 521
- stagflation, p. 536

QUESTIONS FOR REVIEW

1. Name two macroeconomic variables that decline when the economy goes into a recession. Name one macroeconomic variable that rises during a recession.
2. Draw a diagram with aggregate demand, short-run aggregate supply, and long-run aggregate supply. Be careful to label the axes correctly.
3. List and explain the three reasons the aggregate-demand curve is downward sloping.
4. Explain why the long-run aggregate-supply curve is vertical.
5. List and explain the three theories for why the short-run aggregate-supply curve is upward sloping.
6. What might shift the aggregate-demand curve to the left? Use the model of aggregate demand and aggregate supply to trace through the short-run and long-run effects of such a shift on output and the price level.
7. What might shift the aggregate-supply curve to the left? Use the model of aggregate demand and aggregate supply to trace through the short-run and long-run effects of such a shift on output and the price level.

PROBLEMS AND APPLICATIONS

1. Suppose the economy is in a long-run equilibrium.
   a. Draw a diagram to illustrate the state of the economy. Be sure to show aggregate demand, short-run aggregate supply, and long-run aggregate supply.
   b. Now suppose that a stock-market crash causes aggregate demand to fall. Use your diagram to show what happens to output and the price level in the short run. What happens to the unemployment rate?
   c. Use the sticky-wage theory of aggregate supply to explain what will happen to output and the price level in the long run (assuming there is no change in policy). What role does the expected price level play in this adjustment? Be sure to illustrate your analysis in a graph.
2. Explain whether each of the following events will increase, decrease, or have no effect on long-run aggregate supply.
   a. The United States experiences a wave of immigration.
   b. Congress raises the minimum wage to $10 per hour.
   c. Intel invents a new and more powerful computer chip.
   d. A severe hurricane damages factories along the East Coast.
3. Suppose an economy is in long-run equilibrium.
   a. Use the model of aggregate demand and aggregate supply to illustrate the initial equilibrium (call it point A). Be sure to include both short-run and long-run aggregate supply.
b. The central bank raises the money supply by 5 percent. Use your diagram to show what happens to output and the price level as the economy moves from the initial to the new short-run equilibrium (call it point B).

c. Now show the new long-run equilibrium (call it point C). What causes the economy to move from point B to point C?

d. According to the sticky-wage theory of aggregate supply, how do nominal wages at point A compare to nominal wages at point B? How do nominal wages at point A compare to nominal wages at point C?

e. According to the sticky-wage theory of aggregate supply, how do real wages at point A compare to real wages at point B? How do real wages at point A compare to real wages at point C?

f. Judging by the impact of the money supply on nominal and real wages, is this analysis consistent with the proposition that money has real effects in the short run but is neutral in the long run?

4. In 1939, with the U.S. economy not yet fully recovered from the Great Depression, President Roosevelt proclaimed that Thanksgiving would fall a week earlier than usual so that the shopping period before Christmas would be longer. Explain what President Roosevelt might have been trying to achieve, using the model of aggregate demand and aggregate supply.

5. Explain why the following statements are false.
   a. “The aggregate-demand curve slopes downward because it is the horizontal sum of the demand curves for individual goods.”
   b. “The long-run aggregate-supply curve is vertical because economic forces do not affect long-run aggregate supply.”
   c. “If firms adjusted their prices every day, then the short-run aggregate-supply curve would be horizontal.”
   d. “Whenever the economy enters a recession, its long-run aggregate-supply curve shifts to the left.”

6. For each of the three theories for the upward slope of the short-run aggregate-supply curve, carefully explain the following:
   a. How the economy recovers from a recession and returns to its long-run equilibrium without any policy intervention.
   b. What determines the speed of that recovery.

7. Suppose the Fed expands the money supply, but because the public expects this Fed action, it simultaneously raises its expectation of the price level. What will happen to output and the price level in the short run? Compare this result to the outcome if the Fed expanded the money supply but the public didn’t change its expectation of the price level.

8. Suppose that the economy is currently in a recession. If policymakers take no action, how will the economy change over time? Explain in words and using an aggregate-demand/aggregate-supply diagram.

9. The economy begins in long-run equilibrium. Then one day, the president appoints a new chairman of the Federal Reserve. This new chairman is well-known for his view that inflation is not a major problem for an economy.
   a. How would this news affect the price level that people would expect to prevail?
   b. How would this change in the expected price level affect the nominal wage that workers and firms agree to in their new labor contracts?
   c. How would this change in the nominal wage affect the profitability of producing goods and services at any given price level?
   d. How does this change in profitability affect the short-run aggregate-supply curve?
   e. If aggregate demand is held constant, how does this shift in the aggregate-supply curve affect the price level and the quantity of output produced?
   f. Do you think this Fed chairman was a good appointment?

10. Explain whether each of the following events shifts the short-run aggregate-supply curve, the aggregate-demand curve, both, or neither. For each event that does shift a curve, draw a diagram to illustrate the effect on the economy.
   a. Households decide to save a larger share of their income.
   b. Florida orange groves suffer a prolonged period of below-freezing temperatures.
   c. Increased job opportunities overseas cause many people to leave the country.

11. For each of the following events, explain the short-run and long-run effects on output and
the price level, assuming policymakers take no action.

a. The stock market declines sharply, reducing consumers’ wealth.

b. The federal government increases spending on national defense.

c. A technological improvement raises productivity.

d. A recession overseas causes foreigners to buy fewer U.S. goods.

12. Suppose the U.S. economy begins in long-run equilibrium. Concerns about global climate change cause the government to significantly restrict the production of electricity from fossil fuels. Because of this change in policy, foreign investors lose confidence in the economy, and the dollar falls in foreign-exchange markets. Draw a diagram to show the short-run effect of these events, and explain why these changes occur.

13. Suppose firms become very optimistic about future business conditions and invest heavily in new capital equipment.

a. Draw an aggregate-demand/aggregate-supply diagram to show the short-run effect of this optimism on the economy. Label the new levels of prices and real output. Explain in words why the aggregate quantity of output supplied changes.

b. Now use the diagram from part (a) to show the new long-run equilibrium of the economy. (For now, assume there is no change in the long-run aggregate-supply curve.) Explain in words why the aggregate quantity of output demanded changes between the short run and the long run.

c. How might the investment boom affect the long-run aggregate-supply curve? Explain.

14. In economy A, all workers agree in advance on the nominal wages that their employers will pay them. In economy B, half of all workers have these nominal wage contracts, while the other half have indexed employment contracts, so their wages rise and fall automatically with the price level. According to the sticky-wage theory of aggregate supply, which economy has a more steeply sloped short-run aggregate-supply curve? In which economy would a 5 percent increase in the money supply have a larger impact on output? In which economy would it have a larger impact on the price level? Explain.
Imagine that you are a member of the Federal Open Market Committee, the group at the Federal Reserve that sets monetary policy. You observe that the president and Congress have agreed to raise taxes. How should the Fed respond to this change in fiscal policy? Should it expand the money supply, contract the money supply, or leave the money supply the same?

To answer this question, you need to consider the impact of monetary and fiscal policy on the economy. In the preceding chapter, we used the model of aggregate demand and aggregate supply to explain short-run economic fluctuations. We saw that shifts in the aggregate-demand curve or the aggregate-supply curve cause fluctuations in the economy’s overall output of goods and services and its overall level of prices. As we noted in the previous chapter, monetary and fiscal policy can each influence aggregate demand. Thus, a change in one of these policies can lead to short-run fluctuations in output and prices. Policymakers will want to anticipate this effect and, perhaps, adjust the other policy in response.

In this chapter, we examine in more detail how the government’s policy tools influence the position of the aggregate-demand curve. These tools include
monetary policy (the supply of money set by the central bank) and fiscal policy
(the levels of government spending and taxation set by the president and Con-
gress). We have previously discussed the long-run effects of these policies. In
Chapters 17 and 18, we saw how fiscal policy affects saving, investment, and long-
run economic growth. In Chapters 21 and 22, we saw how monetary policy influ-
ences the price level in the long run. We now see how these policy tools can shift
the aggregate-demand curve and, in doing so, affect macroeconomic variables in
the short run.

As we have already learned, many factors influence aggregate demand besides
monetary and fiscal policy. In particular, desired spending by households and
firms determines the overall demand for goods and services. When desired spend-
ing changes, aggregate demand shifts. If policymakers do not respond, such shifts
in aggregate demand cause short-run fluctuations in output and employment. As
a result, monetary and fiscal policymakers sometimes use the policy levers at their
 disposal to try to offset these shifts in aggregate demand and thereby stabilize the
economy. Here we discuss the theory behind these policy actions and some of the
difficulties that arise in using this theory in practice.

HOW MONETARY POLICY INFLUENCES
AGGREGATE DEMAND

The aggregate-demand curve shows the total quantity of goods and services
demanded in the economy for any price level. The preceding chapter discussed
three reasons the aggregate-demand curve slopes downward:

- The wealth effect: A lower price level raises the real value of households’
  money holdings, which are part of their wealth. Higher real wealth stimu-
  lates consumer spending and thus increases the quantity of goods and
  services demanded.
- The interest-rate effect: A lower price level reduces the amount of money
  people want to hold. As people try to lend out their excess money holdings,
  the interest rate falls. The lower interest rate stimulates investment spending
  and thus increases the quantity of goods and services demanded.
- The exchange-rate effect: When a lower price level reduces the interest rate,
  investors move some of their funds overseas in search of higher returns. This
  movement of funds causes the real value of the domestic currency to fall
  in the market for foreign-currency exchange. Domestic goods become less
  expensive relative to foreign goods. This change in the real exchange rate
  stimulates spending on net exports and thus increases the quantity of goods
  and services demanded.

These three effects occur simultaneously to increase the quantity of goods and
services demanded when the price level falls and to decrease it when the price
level rises.

Although all three effects work together to explain the downward slope of the
aggregate-demand curve, they are not of equal importance. Because money hold-
ings are a small part of household wealth, the wealth effect is the least important
of the three. In addition, because exports and imports represent only a small frac-
tion of U.S. GDP, the exchange-rate effect is not large for the U.S. economy. (This
effect is more important for smaller countries, which typically export and import
a higher fraction of their GDP.) For the U.S. economy, the most important reason for the downward slope of the aggregate-demand curve is the interest-rate effect.

To better understand aggregate demand, we now examine the short-run determination of interest rates in more detail. Here we develop the **theory of liquidity preference**. This theory of interest-rate determination will help explain the downward slope of the aggregate-demand curve, as well as how monetary and fiscal policy can shift this curve. By shedding new light on aggregate demand, the theory of liquidity preference expands our understanding of what causes short-run economic fluctuations and what policymakers can potentially do about them.

**The Theory of Liquidity Preference**

In his classic book *The General Theory of Employment, Interest, and Money*, John Maynard Keynes proposed the theory of liquidity preference to explain what factors determine an economy’s interest rate. The theory is, in essence, just an application of supply and demand. According to Keynes, the interest rate adjusts to balance the supply of and demand for money.

You may recall that economists distinguish between two interest rates: The **nominal interest rate** is the interest rate as usually reported, and the **real interest rate** is the interest rate corrected for the effects of inflation. When there is no inflation, the two rates are the same. But when borrowers and lenders expect prices to rise over the course of the loan, they agree to a nominal interest rate that exceeds the real interest rate by the expected rate of inflation. The higher nominal interest rate compensates for the fact that they expect the loan to be repaid in less valuable dollars.

Which interest rate are we now trying to explain with the theory of liquidity preference? The answer is both. In the analysis that follows, we hold constant the expected rate of inflation. This assumption is reasonable for studying the economy in the short run, because expected inflation is typically stable over short periods of time. In this case, nominal and real interest rates differ by a constant. When the nominal interest rate rises or falls, the real interest rate that people expect to earn rises or falls as well. For the rest of this chapter, when we refer to changes in the interest rate, you should envision the real and nominal interest rates moving in the same direction.

Let’s now develop the theory of liquidity preference by considering the supply and demand for money and how each depends on the interest rate.

**Money Supply** The first piece of the theory of liquidity preference is the supply of money. As we first discussed in Chapter 21, the money supply in the U.S. economy is controlled by the Federal Reserve. The Fed alters the money supply primarily by changing the quantity of reserves in the banking system through the purchase and sale of government bonds in open-market operations. When the Fed buys government bonds, the dollars it pays for the bonds are typically deposited in banks, and these dollars are added to bank reserves. When the Fed sells government bonds, the dollars it receives for the bonds are withdrawn from the banking system, and bank reserves fall. These changes in bank reserves, in turn, lead to changes in banks’ ability to make loans and create money. In addition to using open-market operations, the Fed can alter the money supply by changing reserve requirements (the amount of reserves banks must hold against deposits) or the discount rate (the interest rate at which banks can borrow reserves from the Fed).
These details of monetary control, although important for the implementation of Fed policy, are not crucial in this chapter. Our goal here is to examine how changes in the money supply affect the aggregate demand for goods and services. For this purpose, we can ignore the details of how Fed policy is implemented and assume that the Fed controls the money supply directly. In other words, the quantity of money supplied in the economy is fixed at whatever level the Fed decides to set it.

Because the quantity of money supplied is fixed by Fed policy, it does not depend on other economic variables. In particular, it does not depend on the interest rate. Once the Fed has made its policy decision, the quantity of money supplied is the same, regardless of the prevailing interest rate. We represent a fixed money supply with a vertical supply curve, as in Figure 1.

**Money Demand** The second piece of the theory of liquidity preference is the demand for money. As a starting point for understanding money demand, recall that any asset’s liquidity refers to the ease with which that asset can be converted into the economy’s medium of exchange. Because money is the economy’s medium of exchange, it is by definition the most liquid asset available. The liquidity of money explains the demand for it: People choose to hold money instead of other assets that offer higher rates of return because money can be used to buy goods and services.

According to the theory of liquidity preference, the interest rate adjusts to bring the quantity of money supplied and the quantity of money demanded into balance. If the interest rate is above the equilibrium level (such as at $r_1$), the quantity of money people want to hold ($M_d^1$) is less than the quantity the Fed has created, and this surplus of money puts downward pressure on the interest rate. Conversely, if the interest rate is below the equilibrium level (such as at $r_2$), the quantity of money people want to hold ($M_d^2$) is greater than the quantity the Fed has created, and this shortage of money puts upward pressure on the interest rate. Thus, the forces of supply and demand in the market for money push the interest rate toward the equilibrium interest rate, at which people are content holding the quantity of money the Fed has created.
Although many factors determine the quantity of money demanded, the one emphasized by the theory of liquidity preference is the interest rate. The reason is that the interest rate is the opportunity cost of holding money. That is, when you hold wealth as cash in your wallet, instead of as an interest-bearing bond, you lose the interest you could have earned. An increase in the interest rate raises the cost of holding money and, as a result, reduces the quantity of money demanded. A decrease in the interest rate reduces the cost of holding money and raises the quantity demanded. Thus, as shown in Figure 1, the money-demand curve slopes downward.

Equilibrium in the Money Market  According to the theory of liquidity preference, the interest rate adjusts to balance the supply and demand for money. There is one interest rate, called the equilibrium interest rate, at which the quantity of money demanded exactly balances the quantity of money supplied. If the interest rate is at any other level, people will try to adjust their portfolios of assets and, as a result, drive the interest rate toward the equilibrium.

For example, suppose that the interest rate is above the equilibrium level, such as \( r_1 \) in Figure 1. In this case, the quantity of money that people want to hold, \( M_d^1 \), is less than the quantity of money that the Fed has supplied. Those people who are holding the surplus of money will try to get rid of it by buying interest-bearing bonds or by depositing it in an interest-bearing bank account. Because bond issuers and banks prefer to pay lower interest rates, they respond to this surplus of money by lowering the interest rates they offer. As the interest rate falls, people become more willing to hold money until, at the equilibrium interest rate, people are happy to hold exactly the amount of money the Fed has supplied.

Conversely, at interest rates below the equilibrium level, such as \( r_2 \) in Figure 1, the quantity of money that people want to hold, \( M_d^2 \), is greater than the quantity of money that the Fed has supplied. As a result, people try to increase their holdings of money by reducing their holdings of bonds and other interest-bearing assets. As people cut back on their holdings of bonds, bond issuers find that they have to offer higher interest rates to attract buyers. Thus, the interest rate rises and approaches the equilibrium level.

The Downward Slope of the Aggregate-Demand Curve

Having seen how the theory of liquidity preference explains the economy’s equilibrium interest rate, we now consider the theory’s implications for the aggregate demand for goods and services. As a warm-up exercise, let’s begin by using the theory to reexamine a topic we already understand—the interest-rate effect and the downward slope of the aggregate-demand curve. In particular, suppose that the overall level of prices in the economy rises. What happens to the interest rate that balances the supply and demand for money, and how does that change affect the quantity of goods and services demanded?

As we discussed in Chapter 22, the price level is one determinant of the quantity of money demanded. At higher prices, more money is exchanged every time a good or service is sold. As a result, people will choose to hold a larger quantity of money. That is, a higher price level increases the quantity of money demanded for any given interest rate. Thus, an increase in the price level from \( P_1 \) to \( P_2 \) shifts
the money-demand curve to the right from $MD_1$ to $MD_2$, as shown in panel (a) of Figure 2.

Notice how this shift in money demand affects the equilibrium in the money market. For a fixed money supply, the interest rate must rise to balance money supply and money demand. The higher price level has increased the amount of money people want to hold and has shifted the money demand curve to the right. Yet the quantity of money supplied is unchanged, so the interest rate must rise from $r_1$ to $r_2$ to discourage the additional demand.

This increase in the interest rate has ramifications not only for the money market but also for the quantity of goods and services demanded, as shown in panel

---

**FYI**

**Interest Rates in the Long Run and the Short Run**

In an earlier chapter, we said that the interest rate adjusts to balance the supply of loanable funds (national saving) and the demand for loanable funds (desired investment). Here we just said that the interest rate adjusts to balance the supply of and demand for money. Can we reconcile these two theories?

To answer this question, we need to focus on three macroeconomic variables: the economy's output of goods and services, the interest rate, and the price level. According to the classical macroeconomic theory we developed earlier in the book, these variables are determined as follows:

1. **Output** is determined by the supplies of capital and labor and the available production technology for turning capital and labor into output. (We call this the natural rate of output.)
2. For any given level of output, the **interest rate** adjusts to balance the supply and demand for loanable funds.
3. Given output and the interest rate, the **price level** adjusts to balance the supply and demand for money. Changes in the supply of money lead to proportionate changes in the price level.

These are three of the essential propositions of classical economic theory. Most economists believe that these propositions do a good job of describing how the economy works *in the long run.*

Yet these propositions do not hold in the short run. As we discussed in the preceding chapter, many prices are slow to adjust to changes in the money supply; this fact is reflected in a short-run aggregate-supply curve that is upward sloping rather than vertical. As a result, *in the short run,* the overall price level cannot, by itself, move to balance the supply of and demand for money. This stickiness of the price level requires the interest rate to move to bring the money market into equilibrium. These changes in the interest rate, in turn, affect the aggregate demand for goods and services. As aggregate demand fluctuates, the economy's output of goods and services moves away from the level determined by factor supplies and technology.

To think about the operation of the economy in the short run (day to day, week to week, month to month, or quarter to quarter), it is best to keep in mind the following logic:

1. The **price level** is stuck at some level (based on previously formed expectations) and, in the short run, is relatively unresponsive to changing economic conditions.
2. For any given (stuck) price level, the **interest rate** adjusts to balance the supply of and demand for money.
3. The interest rate that balances the money market influences the quantity of goods and services demanded and thus the level of **output**.

Notice that this precisely reverses the order of analysis used to study the economy in the long run.

The two different theories of the interest rate are useful for different purposes. When thinking about the long-run determinants of interest rates, it is best to keep in mind the loanable-funds theory, which highlights the importance of an economy's saving propensities and investment opportunities. By contrast, when thinking about the short-run determinants of interest rates, it is best to keep in mind the liquidity-preference theory, which highlights the importance of monetary policy.
At a higher interest rate, the cost of borrowing and the return to saving are greater. Fewer households choose to borrow to buy a new house, and those who do buy smaller houses, so the demand for residential investment falls. Fewer firms choose to borrow to build new factories and buy new equipment, so business investment falls. Thus, when the price level rises from $P_1$ to $P_2$, increasing money demand from $MD_1$ to $MD_2$ and raising the interest rate from $r_1$ to $r_2$, the quantity of goods and services demanded falls from $Y_1$ to $Y_2$.

This analysis of the interest-rate effect can be summarized in three steps: (1) A higher price level raises money demand. (2) Higher money demand leads to a higher interest rate. (3) A higher interest rate reduces the quantity of goods and services demanded. The same logic works in reverse: A lower price level reduces money demand, which leads to a lower interest rate, and this in turn reduces the quantity of goods and services demanded. The result of this analysis is a negative relationship between the price level and the quantity of goods and services demanded, as illustrated by a downward-sloping aggregate-demand curve.

**Changes in the Money Supply**

So far, we have used the theory of liquidity preference to explain more fully how the total quantity demanded of goods and services in the economy changes as the price level changes. That is, we have examined movements along a downward-sloping aggregate-demand curve. The theory also sheds light, however, on some of the other events that alter the quantity of goods and services
demanded. Whenever the quantity of goods and services demanded changes for any given price level, the aggregate-demand curve shifts.

One important variable that shifts the aggregate-demand curve is monetary policy. To see how monetary policy affects the economy in the short run, suppose that the Fed increases the money supply by buying government bonds in open-market operations. (Why the Fed might do this will become clear later, after we understand the effects of such a move.) Let’s consider how this monetary injection influences the equilibrium interest rate for a given price level. This will tell us what the injection does to the position of the aggregate-demand curve.

As panel (a) of Figure 3 shows, an increase in the money supply shifts the money-supply curve to the right from $MS_1$ to $MS_2$. Because the money-demand curve has not changed, the interest rate falls from $r_1$ to $r_2$ to balance money supply and money demand. That is, the interest rate must fall to induce people to hold the additional money the Fed has created, restoring equilibrium in the money market.

Once again, the interest rate influences the quantity of goods and services demanded, as shown in panel (b) of Figure 3. The lower interest rate reduces the cost of borrowing and the return to saving. Households buy more and larger houses, stimulating the demand for residential investment. Firms spend more on new factories and new equipment, stimulating business investment. As a result, the quantity of goods and services demanded at a given price level, $\bar{P}$, rises from $Y_1$ to $Y_2$. Of course, there is nothing special about $\bar{P}$: The monetary injection raises the quantity of goods and services demanded at every price level. Thus, the entire aggregate-demand curve shifts to the right.

To sum up: When the Fed increases the money supply, it lowers the interest rate and increases the quantity of goods and services demanded for any given price level, shifting...
the aggregate-demand curve to the right. Conversely, when the Fed contracts the money supply, it raises the interest rate and reduces the quantity of goods and services demanded for any given price level, shifting the aggregate-demand curve to the left.

**The Role of Interest-Rate Targets in Fed Policy**

How does the Federal Reserve affect the economy? Our discussion here and earlier in the book has treated the money supply as the Fed’s policy instrument. When the Fed buys government bonds in open-market operations, it increases the money supply and expands aggregate demand. When the Fed sells government bonds in open-market operations, it decreases the money supply and contracts aggregate demand.

Discussions of Fed policy often treat the interest rate, rather than the money supply, as the Fed’s policy instrument. Indeed, in recent years, the Federal Reserve has conducted policy by setting a target for the federal funds rate—the interest rate that banks charge one another for short-term loans. This target is reevaluated every six weeks at meetings of the Federal Open Market Committee (FOMC). The FOMC has chosen to set a target for the federal funds rate, rather than for the money supply, as it has done at times in the past.

There are several, related reasons for the Fed’s decision to use the federal funds as its target. One is that the money supply is hard to measure with sufficient precision. Another is that money demand fluctuates over time. For any given money supply, fluctuations in money demand would lead to fluctuations in interest rates, aggregate demand, and output. By contrast, when the Fed announces a target for the federal funds rate, it essentially accommodates the day-to-day shifts in money demand by adjusting the money supply accordingly.

The Fed’s decision to target an interest rate does not fundamentally alter our analysis of monetary policy. The theory of liquidity preference illustrates an important principle: Monetary policy can be described either in terms of the money supply or in terms of the interest rate. When the FOMC sets a target for the federal funds rate of, say, 6 percent, the Fed’s bond traders are told: “Conduct whatever open-market operations are necessary to ensure that the equilibrium interest rate equals 6 percent.” In other words, when the Fed sets a target for the interest rate, it commits itself to adjusting the money supply to make the equilibrium in the money market hit that target.

As a result, changes in monetary policy can be viewed either in terms of changing the interest rate target or in terms of changing the money supply. When you read in the newspaper that “the Fed has lowered the federal funds rate from 6 to 5 percent,” you should understand that this occurs only because the Fed’s bond traders are doing what it takes to make it happen. To lower the federal funds rate, the Fed’s bond traders buy government bonds, and this purchase increases the money supply and lowers the equilibrium interest rate (just as in Figure 3). Similarly, when the FOMC raises the target for the federal funds rate, the bond traders sell government bonds, and this sale decreases the money supply and raises the equilibrium interest rate.

The lessons from this analysis are simple: Changes in monetary policy aimed at expanding aggregate demand can be described either as increasing the money supply or as lowering the interest rate. Changes in monetary policy aimed at contracting aggregate demand can be described either as decreasing the money supply or as raising the interest rate.
WHY THE FED WATCHES THE STOCK MARKET (AND VICE VERSA)

“Irrational exuberance.” That was how Federal Reserve Chairman Alan Greenspan once described the booming stock market of the late 1990s. He was right that the market was exuberant: Average stock prices increased about fourfold during this decade. And perhaps it was even irrational: In the first few years of the following decade, the stock market took back some of these large gains, as stock prices experienced a pronounced decline, falling by about 40 percent from 2000 to 2003.

How should the Fed respond to stock-market fluctuations? The Fed has no reason to care about stock prices in themselves, but it does have the job of monitoring and responding to developments in the overall economy, and the stock market is a piece of that puzzle. When the stock market booms, households become wealthier, and this increased wealth stimulates consumer spending. In addition, a rise in stock prices makes it more attractive for firms to sell new shares of stock, and this stimulates investment spending. For both reasons, a booming stock market expands the aggregate demand for goods and services.
As we discuss more fully later in the chapter, one of the Fed’s goals is to stabilize aggregate demand, because greater stability in aggregate demand means greater stability in output and the price level. To promote stability, the Fed might respond to a stock-market boom by keeping the money supply lower and interest rates higher than it otherwise would. The contractionary effects of higher interest rates would offset the expansionary effects of higher stock prices. In fact, this analysis does describe Fed behavior: Real interest rates were kept high by historical standards during the “irrationally exuberant” stock-market boom of the late 1990s.

The opposite occurs when the stock market falls. Spending on consumption and investment tends to decline, depressing aggregate demand and pushing the economy toward recession. To stabilize aggregate demand, the Fed would increase the money supply and lower interest rates. And indeed, that is what it typically does. For example, on October 19, 1987, the stock market fell by 22.6 percent—its biggest one-day drop in history. The Fed responded to the market crash by increasing the money supply and lowering interest rates. The federal funds rate fell from 7.7 percent at the beginning of October to 6.6 percent at the end of the month. In part because of the Fed’s quick action, the economy avoided a recession. Similarly, as we discussed in a case study in the preceding chapter, the Fed also reduced interest rates during the stock market declines of 2001 and 2002, but this time monetary policy was not quick enough to avert a recession.

While the Fed keeps an eye on the stock market, stock-market participants also keep an eye on the Fed. Because the Fed can influence interest rates and economic activity, it can alter the value of stocks. For example, when the Fed raises interest rates by reducing the money supply, it makes owning stocks less attractive for two reasons. First, a higher interest rate means that bonds, the alternative to stocks, are earning a higher return. Second, the Fed’s tightening of monetary policy reduces the demand for goods and services, which reduces profits. As a result, stock prices often fall when the Fed raises interest rates.

Quick Quiz Use the theory of liquidity preference to explain how a decrease in the money supply affects the equilibrium interest rate. How does this change in monetary policy affect the aggregate-demand curve?

How Fiscal Policy Influences Aggregate Demand

The government can influence the behavior of the economy not only with monetary policy but also with fiscal policy. Fiscal policy refers to the government’s choices regarding the overall level of government purchases and taxes. Earlier in the book, we examined how fiscal policy influences saving, investment, and growth in the long run. In the short run, however, the primary effect of fiscal policy is on the aggregate demand for goods and services.

Changes in Government Purchases

When policymakers change the money supply or the level of taxes, they shift the aggregate-demand curve indirectly by influencing the spending decisions of firms or households. By contrast, when the government alters its own purchases of goods and services, it shifts the aggregate-demand curve directly.
Suppose, for instance, that the U.S. Department of Defense places a $20 billion order for new fighter planes with Boeing, the large aircraft manufacturer. This order raises the demand for the output produced by Boeing, which induces the company to hire more workers and increase production. Because Boeing is part of the economy, the increase in the demand for Boeing planes means an increase in the total quantity of goods and services demanded at each price level. As a result, the aggregate-demand curve shifts to the right.

By how much does this $20 billion order from the government shift the aggregate-demand curve? At first, one might guess that the aggregate-demand curve shifts to the right by exactly $20 billion. It turns out, however, that this is not the case. There are two macroeconomic effects that make the size of the shift in aggregate demand differ from the change in government purchases. The first—the multiplier effect—suggests the shift in aggregate demand could be larger than $20 billion. The second—the crowding-out effect—suggests the shift in aggregate demand could be smaller than $20 billion. We now discuss each of these effects in turn.

**The Multiplier Effect**

When the government buys $20 billion of goods from Boeing, that purchase has repercussions. The immediate impact of the higher demand from the government is to raise employment and profits at Boeing. Then, as the workers see higher earnings and the firm owners see higher profits, they respond to this increase in income by raising their own spending on consumer goods. As a result, the government purchase from Boeing raises the demand for the products of many other firms in the economy. Because each dollar spent by the government can raise the aggregate demand for goods and services by more than a dollar, government purchases are said to have a multiplier effect on aggregate demand.

This multiplier effect continues even after this first round. When consumer spending rises, the firms that produce these consumer goods hire more people and experience higher profits. Higher earnings and profits stimulate consumer spending once again and so on. Thus, there is positive feedback as higher demand leads to higher income, which in turn leads to even higher demand. Once all these effects are added together, the total impact on the quantity of goods and services demanded can be much larger than the initial impulse from higher government spending.

Figure 4 illustrates the multiplier effect. The increase in government purchases of $20 billion initially shifts the aggregate-demand curve to the right from $AD_1$ to $AD_2$ by exactly $20 billion. But when consumers respond by increasing their spending, the aggregate-demand curve shifts still further to $AD_3$.

This multiplier effect arising from the response of consumer spending can be strengthened by the response of investment to higher levels of demand. For instance, Boeing might respond to the higher demand for planes by deciding to buy more equipment or build another plant. In this case, higher government demand spurs higher demand for investment goods. This positive feedback from demand to investment is sometimes called the investment accelerator.

**A Formula for the Spending Multiplier**

Some simple algebra permits us to derive a formula for the size of the multiplier effect that arises when an increase in government purchases induces increases in
consumer spending. An important number in this formula is the marginal propensity to consume \((MPC)\)—the fraction of extra income that a household consumes rather than saves. For example, suppose that the marginal propensity to consume is \(\frac{3}{4}\). This means that for every extra dollar that a household earns, the household spends \(\$0.75\) (\(\frac{3}{4}\) of the dollar) and saves \(\$0.25\). With an \(MPC\) of \(\frac{3}{4}\), when the workers and owners of Boeing earn \(\$20\) billion from the government contract, they increase their consumer spending by \(\frac{3}{4} \times \$20\) billion, or \(\$15\) billion.

To gauge the impact on aggregate demand of a change in government purchases, we follow the effects step by step. The process begins when the government spends \(\$20\) billion, which implies that national income (earnings and profits) also rises by this amount. This increase in income in turn raises consumer spending by \(MPC \times \$20\) billion, which in turn raises the income for the workers and owners of the firms that produce the consumption goods. This second increase in income again raises consumer spending, this time by \(MPC \times (MPC \times \$20\) billion). These feedback effects go on and on.

To find the total impact on the demand for goods and services, we add up all these effects:

\[
\begin{align*}
\text{Change in government purchases} & = \$20\text{ billion} \\
\text{First change in consumption} & = MPC \times \$20\text{ billion} \\
\text{Second change in consumption} & = MPC^2 \times \$20\text{ billion} \\
\text{Third change in consumption} & = MPC^3 \times \$20\text{ billion} \\
\text{Total change in demand} & = (1 + MPC + MPC^2 + MPC^3 + \ldots) \times \$20\text{ billion}.
\end{align*}
\]
Here “…” represents an infinite number of similar terms. Thus, we can write the multiplier as follows:

\[ \text{Multiplier} = 1 + MPC + MPC^2 + MPC^3 + \ldots. \]

This multiplier tells us the demand for goods and services that each dollar of government purchases generates.

To simplify this equation for the multiplier, recall from math class that this expression is an infinite geometric series. For \( x \) between \(-1\) and \(+1\),

\[ 1 + x + x^2 + x^3 + \ldots = \frac{1}{1 - x}. \]

In our case, \( x = MPC \). Thus,

\[ \text{Multiplier} = \frac{1}{1 - MPC}. \]

For example, if \( MPC \) is \( \frac{3}{4} \), the multiplier is \( \frac{1}{1 - \frac{3}{4}} \), which is 4. In this case, the $20 billion of government spending generates $80 billion of demand for goods and services.

This formula for the multiplier shows an important conclusion: The size of the multiplier depends on the marginal propensity to consume. While an \( MPC \) of \( \frac{3}{4} \) leads to a multiplier of 4, an \( MPC \) of \( \frac{1}{2} \) leads to a multiplier of only 2. Thus, a larger \( MPC \) means a larger multiplier. To see why this is true, remember that the multiplier arises because higher income induces greater spending on consumption. With a larger \( MPC \), consumption responds more to a change in income, and so the multiplier is larger.

**Other Applications of the Multiplier Effect**

Because of the multiplier effect, a dollar of government purchases can generate more than a dollar of aggregate demand. The logic of the multiplier effect, however, is not restricted to changes in government purchases. Instead, it applies to any event that alters spending on any component of GDP—consumption, investment, government purchases, or net exports.

For example, suppose that a recession overseas reduces the demand for U.S. net exports by $10 billion. This reduced spending on U.S. goods and services depresses U.S. national income, which reduces spending by U.S. consumers. If the marginal propensity to consume is \( \frac{3}{4} \) and the multiplier is 4, then the $10 billion fall in net exports means a $40 billion contraction in aggregate demand.

As another example, suppose that a stock-market boom increases households’ wealth and stimulates their spending on goods and services by $20 billion. This extra consumer spending increases national income, which in turn generates even more consumer spending. If the marginal propensity to consume is \( \frac{3}{4} \) and the multiplier is 4, then the initial impulse of $20 billion in consumer spending translates into an $80 billion increase in aggregate demand.

The multiplier is an important concept in macroeconomics because it shows how the economy can amplify the impact of changes in spending. A small initial change in consumption, investment, government purchases, or net exports can end up having a large effect on aggregate demand and, therefore, the economy’s production of goods and services.
THE CROWDING-OUT EFFECT

The multiplier effect seems to suggest that when the government buys $20 billion of planes from Boeing, the resulting expansion in aggregate demand is necessarily larger than $20 billion. Yet another effect is working in the opposite direction. While an increase in government purchases stimulates the aggregate demand for goods and services, it also causes the interest rate to rise, which reduces investment spending and puts downward pressure on aggregate demand. The reduction in aggregate demand that results when a fiscal expansion raises the interest rate is called the crowding-out effect.

To see why crowding out occurs, let’s consider what happens in the money market when the government buys planes from Boeing. As we have discussed, this increase in demand raises the incomes of the workers and owners of this firm (and because of the multiplier effect, of other firms as well). As incomes rise, households plan to buy more goods and services and, as a result, choose to hold more of their wealth in liquid form. That is, the increase in income caused by the fiscal expansion raises the demand for money.

The effect of the increase in money demand is shown in panel (a) of Figure 5. Because the Fed has not changed the money supply, the vertical supply curve remains the same. When the higher level of income shifts the money-demand

Panel (a) shows the money market. When the government increases its purchases of goods and services, the resulting increase in income raises the demand for money from $MD_1$ to $MD_2$, and this causes the equilibrium interest rate to rise from $r_1$ to $r_2$. Panel (b) shows the effects on aggregate demand. The initial impact of the increase in government purchases shifts the aggregate-demand curve from $AD_1$ to $AD_2$. Yet because the interest rate is the cost of borrowing, the increase in the interest rate tends to reduce the quantity of goods and services demanded, particularly for investment goods. This crowding out of investment partially offsets the impact of the fiscal expansion on aggregate demand. In the end, the aggregate-demand curve shifts only to $AD_3$.

FIGURE 5

The Crowding-Out Effect
curve to the right from $MD_1$ to $MD_2$, the interest rate must rise from $r_1$ to $r_2$ to keep supply and demand in balance.

The increase in the interest rate, in turn, reduces the quantity of goods and services demanded. In particular, because borrowing is more expensive, the demand for residential and business investment goods declines. That is, as the increase in government purchases increases the demand for goods and services, it may also crowd out investment. This crowding-out effect partially offsets the impact of government purchases on aggregate demand, as illustrated in panel (b) of Figure 5. The increase in government purchases initially shifts the aggregate-demand curve from $AD_1$ to $AD_2$, but once crowding out takes place, the aggregate-demand curve drops back to $AD_3$.

To sum up: When the government increases its purchases by $20 billion, the aggregate demand for goods and services could rise by more or less than $20 billion depending on the sizes of the multiplier and crowding-out effects. The multiplier effect by itself makes the shift in aggregate demand greater than $20 billion. The crowding-out effect pushes the aggregate-demand curve in the opposite direction and, if large enough, could result in an aggregate-demand shift of less than $20 billion.

**Changes in Taxes**

The other important instrument of fiscal policy, besides the level of government purchases, is the level of taxation. When the government cuts personal income taxes, for instance, it increases households’ take-home pay. Households will save some of this additional income, but they will also spend some of it on consumer goods. Because it increases consumer spending, the tax cut shifts the aggregate-demand curve to the right. Similarly, a tax increase depresses consumer spending and shifts the aggregate-demand curve to the left.

The size of the shift in aggregate demand resulting from a tax change is also affected by the multiplier and crowding-out effects. When the government cuts taxes and stimulates consumer spending, earnings and profits rise, which further stimulates consumer spending. This is the multiplier effect. At the same time, higher income leads to higher money demand, which tends to raise interest rates. Higher interest rates make borrowing more costly, which reduces investment spending. This is the crowding-out effect. Depending on the size of the multiplier and crowding-out effects, the shift in aggregate demand could be larger or smaller than the tax change that causes it.

In addition to the multiplier and crowding-out effects, there is another important determinant of the size of the shift in aggregate demand that results from a tax change: households’ perceptions about whether the tax change is permanent or temporary. For example, suppose that the government announces a tax cut of $1,000 per household. In deciding how much of this $1,000 to spend, households must ask themselves how long this extra income will last. If households expect the tax cut to be permanent, they will view it as adding substantially to their financial resources and, therefore, increase their spending by a large amount. In this case, the tax cut will have a large impact on aggregate demand. By contrast, if households expect the tax change to be temporary, they will view it as adding only slightly to their financial resources and, therefore, will increase their spending by only a small amount. In this case, the tax cut will have a small impact on aggregate demand.

An extreme example of a temporary tax cut was the one announced in 1992. In that year, President George H. W. Bush faced a lingering recession and an upcom-
How Fiscal Policy Might Affect Aggregate Supply

So far, our discussion of fiscal policy has stressed how changes in government purchases and changes in taxes influence the quantity of goods and services demanded. Most economists believe that the short-run macroeconomic effects of fiscal policy work primarily through aggregate demand. Yet fiscal policy can potentially influence the quantity of goods and services supplied as well.

For instance, consider the effects of tax changes on aggregate supply. One of the Ten Principles of Economics in Chapter 1 is that people respond to incentives. When government policymakers cut tax rates, workers get to keep more of each dollar they earn, so they have a greater incentive to work and produce goods and services. If they respond to this incentive, the quantity of goods and services supplied will be greater at each price level, and the aggregate-supply curve will shift to the right.

Some economists, called supply-siders, have argued that the influence of tax cuts on aggregate supply is large. According to some supply-siders, the influence is so large that a cut in tax rates will stimulate enough additional production and income that tax revenue will actually increase. This is certainly a theoretical possibility, but most economists do not consider it the normal case. While the supply-side effects of taxes are important to consider, they are usually not large enough to cause tax revenue to rise when tax rates fall.

Like changes in taxes, changes in government purchases can also potentially affect aggregate supply. Suppose, for instance, that the government increases expenditure on a form of government-provided capital, such as roads. Roads are used by private businesses to make deliveries to their customers; an increase in the quantity of roads increases these businesses’ productivity. Hence, when the government spends more on roads, it increases the quantity of goods and services supplied at any given price level and, thus, shifts the aggregate-supply curve to the right. This effect on aggregate supply is probably more important in the long run than in the short run, however, because it would take some time for the government to build the new roads and put them into use.
THE CASE FOR ACTIVE STABILIZATION POLICY

Let’s return to the question that began this chapter: When the president and Congress raise taxes, how should the Federal Reserve respond? As we have seen, the level of taxation is one determinant of the position of the aggregate-demand curve. When the government raises taxes, aggregate demand will fall, depressing production and employment in the short run. If the Federal Reserve wants to prevent this adverse effect of the fiscal policy, it can expand aggregate demand by increasing the money supply. A monetary expansion would reduce interest rates, stimulate investment spending, and expand aggregate demand. If monetary policy responds appropriately, the combined changes in monetary and fiscal policy could leave the aggregate demand for goods and services unaffected.

This analysis is exactly the sort followed by members of the Federal Open Market Committee. They know that monetary policy is an important determinant of aggregate demand. They also know that there are other important determinants as well, including fiscal policy set by the president and Congress. As a result, the Fed’s Open Market Committee watches the debates over fiscal policy with a keen eye.

This response of monetary policy to the change in fiscal policy is an example of a more general phenomenon: the use of policy instruments to stabilize aggregate demand and, as a result, production and employment. Economic stabilization has been an explicit goal of U.S. policy since the Employment Act of 1946. This act states that “it is the continuing policy and responsibility of the federal government to . . . promote full employment and production.” In essence, the government has chosen to hold itself accountable for short-run macroeconomic performance.

The Employment Act has two implications. The first, more modest, implication is that the government should avoid being a cause of economic fluctuations. Thus, most economists advise against large and sudden changes in monetary and fiscal policy, for such changes are likely to cause fluctuations in aggregate demand. Moreover, when large changes do occur, it is important that monetary and fiscal policymakers be aware of and respond to the others’ actions.

The second, more ambitious, implication of the Employment Act is that the government should respond to changes in the private economy to stabilize aggregate demand. The act was passed not long after the publication of Keynes’s *The General Theory of Employment, Interest, and Money*, which has been one of the most influential books ever written about economics. In it, Keynes emphasized the key role of aggregate demand in explaining short-run economic fluctuations. Keynes claimed that the government should actively stimulate aggregate demand when aggregate demand appeared insufficient to maintain production at its full-employment level.

Keynes (and his many followers) argued that aggregate demand fluctuates because of largely irrational waves of pessimism and optimism. He used the term “animal spirits” to refer to these arbitrary changes in attitude. When pessimism reigns, households reduce consumption spending, and firms reduce investment spending. The result is reduced aggregate demand, lower production, and higher unemployment. Conversely, when optimism reigns, households and firms increase spending. The result is higher aggregate demand, higher production, and inflationary pressure. Notice that these changes in attitude are, to some extent, self-fulfilling.

In principle, the government can adjust its monetary and fiscal policy in response to these waves of optimism and pessimism and, thereby, stabilize the
economy. For example, when people are excessively pessimistic, the Fed can expand the money supply to lower interest rates and expand aggregate demand. When they are excessively optimistic, it can contract the money supply to raise interest rates and dampen aggregate demand. Former Fed Chairman William McChesney Martin described this view of monetary policy very simply: “The Federal Reserve’s job is to take away the punch bowl just as the party gets going.”

KEYNESIANS IN THE WHITE HOUSE

When a reporter in 1961 asked President John F. Kennedy why he advocated a tax cut, Kennedy replied, “To stimulate the economy. Don’t you remember your Economics 101?” Kennedy’s policy was, in fact, based on the analysis of fiscal policy we have developed in this chapter. His goal was to enact a tax cut, which would raise consumer spending, expand aggregate demand, and increase the economy’s production and employment.

In choosing this policy, Kennedy was relying on his team of economic advisers. This team included such prominent economists as James Tobin and Robert Solow, who later would win Nobel Prizes for their contributions to economics. As students in the 1940s, these economists had closely studied John Maynard Keynes’s *General Theory*, which then was only a few years old. When the Kennedy advisers proposed cutting taxes, they were putting Keynes’s ideas into action.

Although tax changes can have a potent influence on aggregate demand, they have other effects as well. In particular, by changing the incentives that people face, taxes can alter the aggregate supply of goods and services. Part of the Kennedy proposal was an investment tax credit, which gives a tax break to firms that invest in new capital. Higher investment would not only stimulate aggregate demand immediately but also increase the economy’s productive capacity over time. Thus, the short-run goal of increasing production through higher aggregate demand was coupled with a long-run goal of increasing production through higher aggregate supply. And indeed, when the tax cut Kennedy proposed was finally enacted in 1964, it helped usher in a period of robust economic growth.

Since the 1964 tax cut, policymakers have from time to time used fiscal policy as a tool for controlling aggregate demand. For example, when President George W. Bush moved into the Oval Office in 2001, he faced an economy that was heading into recession. One of his first policy initiatives was a substantial and permanent tax cut. Bush explained, “The best way to increase demand for goods and services is to let people keep more of their own money. And when somebody meets that demand by additional production, somebody is more likely to find a job.” The explanation sounds like it could have come from an exam for Econ 101.

THE CASE AGAINST ACTIVE STABILIZATION POLICY

Some economists argue that the government should avoid active use of monetary and fiscal policy to try to stabilize the economy. They claim that these policy instruments should be set to achieve long-run goals, such as rapid economic growth and low inflation, and that the economy should be left to deal with short-run fluctuations on its own. Although these economists may admit that monetary and fiscal policy can stabilize the economy in theory, they doubt whether it can do so in practice.
The primary argument against active monetary and fiscal policy is that these policies affect the economy with a long lag. As we have seen, monetary policy works by changing interest rates, which in turn influence investment spending. But many firms make investment plans far in advance. Thus, most economists believe that it takes at least six months for changes in monetary policy to have much effect on output and employment. Moreover, once these effects occur, they can last for several years. Critics of stabilization policy argue that because of this lag, the Fed should not try to fine-tune the economy. They claim that the Fed often reacts too late to changing economic conditions and, as a result, ends up being a cause of rather than a cure for economic fluctuations. These critics advocate a passive monetary policy, such as slow and steady growth in the money supply.

Fiscal policy also works with a lag, but unlike the lag in monetary policy, the lag in fiscal policy is largely attributable to the political process. In the United States, most changes in government spending and taxes must go through congressional committees in both the House and the Senate, be passed by both legislative bodies, and then be signed by the president. Completing this process can take months and, in some cases, years. By the time the change in fiscal policy is passed and ready to implement, the condition of the economy may well have changed.

These lags in monetary and fiscal policy are a problem in part because economic forecasting is so imprecise. If forecasters could accurately predict the condition of the economy a year in advance, then monetary and fiscal policymakers could look ahead when making policy decisions. In this case, policymakers could stabilize the economy despite the lags they face. In practice, however, major recessions and depressions arrive without much advance warning. The best policymakers can do at any time is to respond to economic changes as they occur.
AUTOMATIC STABILIZERS

All economists—both advocates and critics of stabilization policy—agree that the lags in implementation render policy less useful as a tool for short-run stabilization. The economy would be more stable, therefore, if policymakers could find a way to avoid some of these lags. In fact, they have. **Automatic stabilizers** are changes in fiscal policy that stimulate aggregate demand when the economy goes into a recession without policymakers having to take any deliberate action.

The most important automatic stabilizer is the tax system. When the economy goes into a recession, the amount of taxes collected by the government falls automatically because almost all taxes are closely tied to economic activity. The personal income tax depends on households’ incomes, the payroll tax depends on workers’ earnings, and the corporate income tax depends on firms’ profits. Because incomes, earnings, and profits all fall in a recession, the government’s tax revenue falls as well. This automatic tax cut stimulates aggregate demand and, thereby, reduces the magnitude of economic fluctuations.

Government spending also acts as an automatic stabilizer. In particular, when the economy goes into a recession and workers are laid off, more people apply for unemployment insurance benefits, welfare benefits, and other forms of income support. This automatic increase in government spending stimulates aggregate demand at exactly the time when aggregate demand is insufficient to maintain full employment. Indeed, when the unemployment insurance system was first enacted in the 1930s, economists who advocated this policy did so in part because of its power as an automatic stabilizer.

The automatic stabilizers in the U.S. economy are not sufficiently strong to prevent recessions completely. Nonetheless, without these automatic stabilizers, output and employment would probably be more volatile than they are. For this reason, many economists oppose a constitutional amendment that would require the federal government always to run a balanced budget, as some politicians have proposed. When the economy goes into a recession, taxes fall, government spending rises, and the government’s budget moves toward deficit. If the government faced a strict balanced-budget rule, it would be forced to look for ways to raise taxes or cut spending in a recession. In other words, a strict balanced-budget rule would eliminate the automatic stabilizers inherent in our current system of taxes and government spending.

**QUICK QUIZ**  Suppose a wave of negative “animal spirits” overruns the economy, and people become pessimistic about the future. What happens to aggregate demand? If the Fed wants to stabilize aggregate demand, how should it alter the money supply? If it does this, what happens to the interest rate? Why might the Fed choose not to respond in this way?

**CONCLUSION**

Before policymakers make any change in policy, they need to consider all the effects of their decisions. Earlier in the book, we examined classical models of the economy, which describe the long-run effects of monetary and fiscal policy. There we saw how fiscal policy influences saving, investment, and long-run growth and how monetary policy influences the price level and the inflation rate.
In this chapter, we examined the short-run effects of monetary and fiscal policy. We saw how these policy instruments can change the aggregate demand for goods and services and alter the economy’s production and employment in the short run. When Congress reduces government spending to balance the budget, it needs to consider both the long-run effects on saving and growth and the short-run effects on aggregate demand and employment. When the Fed reduces the growth rate of the money supply, it must take into account the long-run effect on inflation as well as the short-run effect on production. In all parts of government, policymakers must keep in mind both long-run and short-run goals.

SUMMARY

- In developing a theory of short-run economic fluctuations, Keynes proposed the theory of liquidity preference to explain the determinants of the interest rate. According to this theory, the interest rate adjusts to balance the supply and demand for money.

- An increase in the price level raises money demand and increases the interest rate that brings the money market into equilibrium. Because the interest rate represents the cost of borrowing, a higher interest rate reduces investment and, thereby, the quantity of goods and services demanded. The downward-sloping aggregate-demand curve expresses this negative relationship between the price level and the quantity demanded.

- Policymakers can influence aggregate demand with monetary policy. An increase in the money supply reduces the equilibrium interest rate for any given price level. Because a lower interest rate stimulates investment spending, the aggregate-demand curve shifts to the right. Conversely, a decrease in the money supply raises the equilibrium interest rate for any given price level and shifts the aggregate-demand curve to the left.

- Policymakers can also influence aggregate demand with fiscal policy. An increase in government purchases or a cut in taxes shifts the aggregate-demand curve to the right. A decrease in government purchases or an increase in taxes shifts the aggregate-demand curve to the left.

- When the government alters spending or taxes, the resulting shift in aggregate demand can be larger or smaller than the fiscal change. The multiplier effect tends to amplify the effects of fiscal policy on aggregate demand. The crowding-out effect tends to dampen the effects of fiscal policy on aggregate demand.

- Because monetary and fiscal policy can influence aggregate demand, the government sometimes uses these policy instruments in an attempt to stabilize the economy. Economists disagree about how active the government should be in this effort. According to advocates of active stabilization policy, changes in attitudes by households and firms shift aggregate demand; if the government does not respond, the result is undesirable and unnecessary fluctuations in output and employment. According to critics of active stabilization policy, monetary and fiscal policy work with such long lags that attempts at stabilizing the economy often end up being destabilizing.

KEY CONCEPTS

theory of liquidity preference, p. 547
fiscal policy, p. 555
multiplier effect, p. 556
crowding-out effect, p. 559
automatic stabilizers, p. 565
CHAPTER 24  THE INFLUENCE OF MONETARY AND FISCAL POLICY ON AGGREGATE DEMAND  567

QUESTIONS FOR REVIEW

1. What is the theory of liquidity preference? How does it help explain the downward slope of the aggregate-demand curve?
2. Use the theory of liquidity preference to explain how a decrease in the money supply affects the aggregate-demand curve.
3. The government spends $3 billion to buy police cars. Explain why aggregate demand might increase by more than $3 billion. Explain why aggregate demand might increase by less than $3 billion.
4. Suppose that survey measures of consumer confidence indicate a wave of pessimism is sweeping the country. If policymakers do nothing, what will happen to aggregate demand? What should the Fed do if it wants to stabilize aggregate demand? If the Fed does nothing, what might Congress do to stabilize aggregate demand?
5. Give an example of a government policy that acts as an automatic stabilizer. Explain why the policy has this effect.

PROBLEMS AND APPLICATIONS

1. Explain how each of the following developments would affect the supply of money, the demand for money, and the interest rate. Illustrate your answers with diagrams.
   a. The Fed’s bond traders buy bonds in open-market operations.
   b. An increase in credit-card availability reduces the cash people hold.
   c. The Federal Reserve reduces banks’ reserve requirements.
   d. Households decide to hold more money to use for holiday shopping.
   e. A wave of optimism boosts business investment and expands aggregate demand.
2. The Federal Reserve expands the money supply by 5 percent.
   a. Use the theory of liquidity preference to illustrate in a graph the impact of this policy on the interest rate.
   b. Use the model of aggregate demand and aggregate supply to illustrate the impact of this change in the interest rate on output and the price level in the short run.
   c. When the economy makes the transition from its short-run equilibrium to its long-run equilibrium, what will happen to the price level?
   d. How will this change in the price level affect the demand for money and the equilibrium interest rate?
   e. Is this analysis consistent with the proposition that money has real effects in the short run but is neutral in the long run?
3. Suppose banks install automatic teller machines on every block and, by making cash readily available, reduce the amount of money people want to hold.
   a. Assume the Fed does not change the money supply. According to the theory of liquidity preference, what happens to the interest rate? What happens to aggregate demand?
   b. If the Fed wants to stabilize aggregate demand, how should it respond?
4. Consider two policies—a tax cut that will last for only 1 year and a tax cut that is expected to be permanent. Which policy will stimulate greater spending by consumers? Which policy will have the greater impact on aggregate demand? Explain.
5. The economy is in a recession with high unemployment and low output.
   a. Draw a graph of aggregate demand on every block and, by making cash readily available, reduce the amount of money people want to hold.
   b. Assume the Fed does not change the money supply. According to the theory of liquidity preference, what happens to the interest rate? What happens to aggregate demand?
   c. If the Fed wants to stabilize aggregate demand, how should it respond?
4. Consider two policies—a tax cut that will last for only 1 year and a tax cut that is expected to be permanent. Which policy will stimulate greater spending by consumers? Which policy will have the greater impact on aggregate demand? Explain.
5. The economy is in a recession with high unemployment and low output.
   a. Draw a graph of aggregate demand and aggregate supply to illustrate the current situation. Be sure to include the aggregate-demand curve, the short-run aggregate-supply curve, and the long-run aggregate-supply curve.
   b. Identify an open-market operation that would restore the economy to its natural rate.
   c. Draw a graph of the money market to illustrate the effect of this open-market operation. Show the resulting change in the interest rate.
   d. Draw a graph similar to the one in part (a) to show the effect of the open-market operation on output and the price level. Explain in
no crowding out. The marginal propensity to consume is \( \frac{4}{5} \), and the price level is completely fixed in the short run.

a. In what direction and by how much would government spending need to change to close the recessionary gap? Explain your thinking.

b. In what direction and by how much would taxes need to change to close the gap? Explain.

c. If the central bank were to hold the money supply, rather than the interest rate, constant in response to the change in fiscal policy, would your answers to the previous questions be larger, smaller, or the same? Explain.

d. If policymakers in this economy wanted to close the recessionary gap without increasing the government’s budget deficit, what are two ways they can accomplish this goal?

10. Suppose government spending increases. Would the effect on aggregate demand be larger if the Federal Reserve took no action in response or if the Fed were committed to maintaining a fixed interest rate? Explain.

11. In which of the following circumstances is expansionary fiscal policy more likely to lead to a short-run increase in investment? Explain.

a. When the investment accelerator is large or when it is small?

b. When the interest sensitivity of investment is large or when it is small?

12. For various reasons, fiscal policy changes automatically when output and employment fluctuate.

a. Explain why tax revenue changes when the economy goes into a recession.

b. Explain why government spending changes when the economy goes into a recession.

c. If the government were to operate under a strict balanced-budget rule, what would it have to do in a recession? Would that make the recession more or less severe?

13. Some members of Congress have proposed a law that would make price stability the sole goal of monetary policy. Suppose such a law were passed.

a. How would the Fed respond to an event that contracted aggregate demand?

b. How would the Fed respond to an event that caused an adverse shift in short-run aggregate supply?

In each case, is there another monetary policy that would lead to greater stability in output?
absolute advantage  the ability to produce a good using fewer inputs than another producer
accounting profit  total revenue minus total explicit cost
aggregate-demand curve  a curve that shows the quantity of goods and services that households, firms, the government, and customers abroad want to buy at each price level
aggregate-supply curve  a curve that shows the quantity of goods and services that firms choose to produce and sell at each price level
automatic stabilizers  changes in fiscal policy that stimulate aggregate demand when the economy goes into a recession without policymakers having to take any deliberate action
average fixed cost  fixed cost divided by the quantity of output
average revenue  total revenue divided by the quantity sold
average total cost  total cost divided by the quantity of output
average variable cost  variable cost divided by the quantity of output

bond  a certificate of indebtedness
budget deficit  a shortfall of tax revenue from government spending
budget surplus  an excess of tax revenue over government spending
business cycle  fluctuations in economic activity, such as employment and production

catch-up effect  the property whereby countries that start off poor tend to grow more rapidly than countries that start off rich
central bank  an institution designed to oversee the banking system and regulate the quantity of money in the economy
circular-flow diagram  a visual model of the economy that shows how dollars flow through markets among households and firms
classical dichotomy  the theoretical separation of nominal and real variables
Coase theorem  the proposition that if private parties can bargain without cost over the allocation of resources, they can solve the problem of externalities on their own
collective bargaining  the process by which unions and firms agree on the terms of employment
commodity money  money that takes the form of a commodity with intrinsic value
common resources  goods that are rival in consumption but not excludable
comparative advantage  the ability to produce a good at a lower opportunity cost than another producer
competitive market  a market with many buyers and sellers trading identical products so that each buyer and seller is a price taker
complements  two goods for which an increase in the price of one leads to a decrease in the demand for the other
compounding  the accumulation of a sum of money in, say, a bank account, where the interest earned remains in the account to earn additional interest in the future
constant returns to scale  the property whereby long-run average total cost stays the same as the quantity of output changes
consumer price index (CPI)  a measure of the overall cost of the goods and services bought by a typical consumer
consumer surplus  the amount a buyer is willing to pay for a good minus the amount the buyer actually pays for it
consumption  spending by households on goods and services, with the exception of purchases of new housing
corrective tax  a tax designed to induce private decision makers to take account of the social costs that arise from a negative externality
cost  the value of everything a seller must give up to produce a good
cost–benefit analysis  a study that compares the costs and benefits to society of providing a public good
cross-price elasticity of demand  a measure of how much the quantity demanded of one good responds to a change in the price of another good, computed as the percentage change in quantity demanded of the first good divided by the percentage change in the price of the second good
crowding out  a decrease in investment that results from government borrowing
crowding-out effect  the offset in aggregate demand that results when expansionary fiscal policy raises the interest rate and thereby reduces investment spending
currency the paper bills and coins in the hands of the public
cyclical unemployment the deviation of unemployment from its natural rate
deadweight loss the fall in total surplus that results from a market distortion, such as a tax
demand curve a graph of the relationship between the price of a good and the quantity demanded
demand deposits balances in bank accounts that depositors can access on demand by writing a check
demand schedule a table that shows the relationship between the price of a good and the quantity demanded
depression a severe recession
diminishing marginal product the property whereby the marginal product of an input declines as the quantity of the input increases
diminishing returns the property whereby the benefit from an extra unit of an input declines as the quantity of the input increases
discount rate the interest rate on the loans that the Fed makes to banks
discouraged workers individuals who would like to work but have given up looking for a job
diseconomies of scale the property whereby long-run average total cost rises as the quantity of output increases
diversification the reduction of risk achieved by replacing a single risk with a large number of smaller, unrelated risks
economic profit total revenue minus total cost, including both explicit and implicit costs
economics the study of how society manages its scarce resources
economies of scale the property whereby long-run average total cost falls as the quantity of output increases
efficiency the property of society getting the most it can from its scarce resources
efficiency wages above-equilibrium wages paid by firms to increase worker productivity
efficient markets hypothesis the theory that asset prices reflect all publicly available information about the value of an asset
efficient scale the quantity of output that minimizes average total cost
elasticity a measure of the responsiveness of quantity demanded or quantity supplied to one of its determinants
equality the property of distributing economic prosperity uniformly among the members of society
equilibrium a situation in which the market price has reached the level at which quantity supplied equals quantity demanded
equilibrium price the price that balances quantity supplied and quantity demanded
equilibrium quantity the quantity supplied and the quantity demanded at the equilibrium price
excludability the property of a good whereby a person can be prevented from using it
explicit costs input costs that require an outlay of money by the firm
exports goods produced domestically and sold abroad
externality the uncompensated impact of one person’s actions on the well-being of a bystander
federal funds rate the interest rate at which banks make overnight loans to one another
Federal Reserve (Fed) the central bank of the United States
fiat money money without intrinsic value that is used as money because of government decree
finance the field that studies how people make decisions regarding the allocation of resources over time and the handling of risk
financial intermediaries financial institutions through which savers can indirectly provide funds to borrowers
financial markets financial institutions through which savers can directly provide funds to borrowers
financial system the group of institutions in the economy that help to match one person’s saving with another person’s investment
firm-specific risk risk that affects only a single company
fiscal policy the setting of the level of government spending and taxation by government policymakers
Fisher effect the one-for-one adjustment of the nominal interest rate to the inflation rate
fixed costs costs that do not vary with the quantity of output produced
fractional-reserve banking a banking system in which banks hold only a fraction of deposits as reserves
free rider a person who receives the benefit of a good but avoids paying for it
frictional unemployment unemployment that results because it takes time for workers to search for the jobs that best suit their tastes and skills
fundamental analysis the study of a company’s accounting statements and future prospects to determine its value
future value the amount of money in the future that an amount of money today will yield, given prevailing interest rates
GDP deflator a measure of the price level calculated as the ratio of nominal GDP to real GDP times 100
government purchases spending on goods and services by local, state, and federal governments
**gross domestic product (GDP)**  the market value of all final goods and services produced within a country in a given period of time

**human capital**  the knowledge and skills that workers acquire through education, training, and experience

**implicit costs**  input costs that do not require an outlay of money by the firm

**imports**  goods produced abroad and sold domestically

**incentive**  something that induces a person to act

**income elasticity of demand**  a measure of how much the quantity demanded of a good responds to a change in consumers’ income, computed as the percentage change in quantity demanded divided by the percentage change in income

**indexation**  the automatic correction by law or contract of a dollar amount for the effects of inflation

**inferior good**  a good for which, other things equal, an increase in income leads to a decrease in demand

**inflation**  an increase in the overall level of prices in the economy

**inflation rate**  the percentage change in the price index from the preceding period

**inflation tax**  the revenue the government raises by creating money

**informational efficiency**  the description of asset prices that rationally reflect all available information

**internalizing the externality**  altering incentives so that people take account of the external effects of their actions

**investment**  spending on capital equipment, inventories, and structures, including household purchases of new housing

**job search**  the process by which workers find appropriate jobs given their tastes and skills

**labor force**  the total number of workers, including both the employed and the unemployed

**labor-force participation rate**  the percentage of the adult population that is in the labor force

**law of demand**  the claim that, other things equal, the quantity demanded of a good falls when the price of the good rises

**law of supply**  the claim that, other things equal, the quantity supplied of a good rises when the price of the good rises

**law of supply and demand**  the claim that the price of any good adjusts to bring the quantity supplied and the quantity demanded for that good into balance

**liquidity**  the ease with which an asset can be converted into the economy’s medium of exchange

**macroeconomics**  the study of economy-wide phenomena, including inflation, unemployment, and economic growth

**marginal changes**  small incremental adjustments to a plan of action

**marginal cost**  the increase in total cost that arises from an extra unit of production

**marginal product**  the increase in output that arises from an additional unit of input

**marginal revenue**  the change in total revenue from an additional unit sold

**market**  a group of buyers and sellers of a particular good or service

**market economy**  an economy that allocates resources through the decentralized decisions of many firms and households as they interact in markets for goods and services

**market failure**  a situation in which a market left on its own fails to allocate resources efficiently

**market for loanable funds**  the market in which those who want to save supply funds and those who want to borrow to invest demand funds

**market power**  the ability of a single economic actor (or small group of actors) to have a substantial influence on market prices

**market risk**  risk that affects all companies in the stock market

**medium of exchange**  an item that buyers give to sellers when they want to purchase goods and services

**menu costs**  the costs of changing prices

**microeconomics**  the study of how households and firms make decisions and how they interact in markets

**model of aggregate demand and aggregate supply**  the model that most economists use to explain short-run fluctuations in economic activity around its long-run trend

**monetary neutrality**  the proposition that changes in the money supply do not affect real variables

**monetary policy**  the setting of the money supply by policymakers in the central bank

**money**  the set of assets in an economy that people regularly use to buy goods and services from other people

**money multiplier**  the amount of money the banking system generates with each dollar of reserves

**money supply**  the quantity of money available in the economy

**monopoly**  a firm that is the sole seller of a product without close substitutes

**multiplier effect**  the additional shifts in aggregate demand that result when expansionary fiscal policy increases income and thereby increases consumer spending

**mutual fund**  an institution that sells shares to the public and uses the proceeds to buy a portfolio of stocks and bonds
national saving (saving) the total income in the economy that remains after paying for consumption and government purchases
natural monopoly a monopoly that arises because a single firm can supply a good or service to an entire market at a smaller cost than could two or more firms
natural rate of output the production of goods and services that an economy achieves in the long run when unemployment is at its normal rate
natural rate of unemployment the normal rate of unemployment around which the unemployment rate fluctuates
natural resources the inputs into the production of goods and services that are provided by nature, such as land, rivers, and mineral deposits
net exports spending on domestically produced goods by foreigners (exports) minus spending on foreign goods by domestic residents (imports)
nominal GDP the production of goods and services valued at current prices
nominal interest rate the interest rate as usually reported without a correction for the effects of inflation
nominal variables variables measured in monetary units
normal good a good for which, other things equal, an increase in income leads to an increase in demand
normative statements claims that attempt to prescribe how the world should be
opportunity cost whatever must be given up to obtain some item
physical capital the stock of equipment and structures that are used to produce goods and services
positive statements claims that attempt to describe the world as it is
present value the amount of money today that would be needed, using prevailing interest rates, to produce a given future amount of money
price ceiling a legal maximum on the price at which a good can be sold
price discrimination the business practice of selling the same good at different prices to different customers
price elasticity of demand a measure of how much the quantity demanded of a good responds to a change in the price of that good, computed as the percentage change in quantity demanded divided by the percentage change in price
price elasticity of supply a measure of how much the quantity supplied of a good responds to a change in the price of that good, computed as the percentage change in quantity supplied divided by the percentage change in price
price floor a legal minimum on the price at which a good can be sold
private goods goods that are both excludable and rival in consumption
private saving the income that households have left after paying for taxes and consumption
producer price index a measure of the cost of a basket of goods and services bought by firms
producer surplus the amount a seller is paid for a good above the seller’s cost of providing it
production function the relationship between quantity of inputs used to make a good and the quantity of output of that good
production possibilities frontier a graph that shows the combinations of output that the economy can possibly produce given the available factors of production and the available production technology
productivity the quantity of goods and services produced from each unit of labor input
profit total revenue minus total cost
property rights the ability of an individual to own and exercise control over scarce resources
public goods goods that are neither excludable nor rival in consumption
public saving the tax revenue that the government has left after paying for its spending
quantity demanded the amount of a good that buyers are willing and able to purchase
quantity equation the equation $M \times V = P \times Y$ relates the quantity of money, the velocity of money, and the dollar value of the economy’s output of goods and services
quantity supplied the amount of a good that sellers are willing and able to sell
quantity theory of money a theory asserting that the quantity of money available determines the price level and that the growth rate in the quantity of money available determines the inflation rate
random walk the path of a variable whose changes are impossible to predict
rational people people who systematically and purposefully do the best they can to achieve their objectives
real GDP the production of goods and services valued at constant prices
real interest rate the interest rate corrected for the effects of inflation
real variables variables measured in physical units
recession a period of declining real incomes and rising unemployment
reserve ratio the fraction of deposits that banks hold as reserves
reserve requirements regulations on the minimum amount of reserves that banks must hold against deposits
reserves deposits that banks have received but have not loaned out
risk aversion  a dislike of uncertainty
rivalry in consumption  the property of a good whereby one person's use diminishes other people's use

scarcity  the limited nature of society's resources
shoeleather costs  the resources wasted when inflation encourages people to reduce their money holdings
shortage  a situation in which quantity demanded is greater than quantity supplied
stagflation  a period of falling output and rising prices
stock  a claim to partial ownership in a firm
store of value  an item that people can use to transfer purchasing power from the present to the future
strike  the organized withdrawal of labor from a firm by a union
structural unemployment  unemployment that results because the number of jobs available in some labor markets is insufficient to provide a job for everyone who wants one
substitutes  two goods for which an increase in the price of one leads to an increase in the demand for the other
sunk cost  a cost that has already been committed and cannot be recovered

supply curve  a graph of the relationship between the price of a good and the quantity supplied
supply schedule  a table that shows the relationship between the price of a good and the quantity supplied
surplus  a situation in which quantity supplied is greater than quantity demanded
tariff  a tax on goods produced abroad and sold domestically
tax incidence  the manner in which the burden of a tax is shared among participants in a market

technological knowledge  society's understanding of the best ways to produce goods and services
theory of liquidity preference  Keynes's theory that the interest rate adjusts to bring money supply and money demand into balance

total cost  the market value of the inputs a firm uses in production
total revenue (for firm)  the amount a firm receives for the sale of its output

total revenue (in a market)  the amount paid by buyers and received by sellers of a good, computed as the price of the good times the quantity sold

Tragedy of the Commons  a parable that illustrates why common resources are used more than is desirable from the standpoint of society as a whole

transaction costs  the costs that parties incur in the process of agreeing to and following through on a bargain

unemployment insurance  a government program that partially protects workers' incomes when they become unemployed
unemployment rate  the percentage of the labor force that is unemployed
union  a worker association that bargains with employers over wages, benefits, and working conditions
unit of account  the yardstick people use to post prices and record debts

variable costs  costs that vary with the quantity of output produced
velocity of money  the rate at which money changes hands

welfare economics  the study of how the allocation of resources affects economic well-being
willingness to pay  the maximum amount that a buyer will pay for a good
world price  the price of a good that prevails in the world market for that good
Index

Note: Page numbers in **boldface** refer to pages where key terms are defined.

Absolute advantage, 54
Absolute value, 91
Accidents, associated with driving, 211
Account, unit of, 459
Accountants, vs. economists and opportunity costs, 245
Accounting, 397
Accounting profit, 246
Adverse selection insurance and, 419
Aeppel, Timothy, 350–351
Aggregate demand. See also Aggregate-demand curve; Model of aggregate demand and aggregate supply
automatic stabilizers, 565
changes in taxes and, 560–561
crowding-out effect, 559–560
fiscal policy influences on, 555–561
Great Depression and World War II, 533–534
monetary policy influence on, 546–555
multiplier effect, 556–558
recession of 2001 and, 535
shifts in, and economic fluctuations, 530–535
stabilization policy, 562–564
tax rebate fiscal stimulus package, 518
taxation, 443–445
taxation in the United States, 442–445
theory of liquidity preference, 547–553
Aggregate-demand curve, 513. See also Model of aggregate demand and aggregate supply
downward slope of, 514–517, 549–551
economic fluctuations caused by shifts in, 530–535
exchange-rate effect, 515–516
interest-rate effect, 515
money supply and, 551–553
shifts in, 517–519, 530–535
summary of, 520
wealth effect and, 515
 Aggregate-supply curve, 513, 518–529. See also Model of aggregate demand and aggregate supply
accommodating shifts in, 537–538
economic fluctuations caused by shifts in, 535–539
fiscal policy and, 561
misperceptions theory, 527
natural rate of output, 521
oil industry and shifts in, 538–539
shifts in, 521–523, 535–539
in short-run, 523–529
stagflation, 536
sticky-price theory, 526
sticky-wage theory, 525–526
summary of, 527–528, 529
vertical in long run, 520–521
Agriculture
price controls in Venezuela, 122–123
supply, demand and elasticity, 103–105
Airline industry, price discrimination in, 305–306
Alkire, Caroline, 37
Almanac, 418
Antipoverty programs fighting, as public good, 229–230
Antitrust laws, 308–309
Arbitrage, 303–304
Argentina
GDP of, 367
underground economy in, 336
Assumptions, 23
AT&T, 308
Auerbach, Alan, 170
Australia, underground economy in, 336
taxation, 443–445
taxation in the United States, 442–445
Austrian hyperinflation in, 488–489
Automatic stabilizers, 565
Automobile industry
safety laws, 7–8
Average fixed cost, 252, 259
Average fixed cost curve, 253, 255
Average revenue, 267
Average total cost, 252, 259
related to marginal cost, 254
in short-run vs. long-run, 256–257
Average total cost (ATC) curve, 253, 254, 255
Average variable cost, 252, 259
Average variable cost (AVC) curve, 253, 255
Bailey, Ronald, 374–375
Balanced budget, 405
Ball, Laurence, 574
Bangladesh
GDP of, 367
living standard in, 339
Banks
central, 464. See also European Central Bank (ECB); Federal Reserve System (Fed)
federal funds rate, 473–474
Fed’s tools for monetary control, 469–472
as financial intermediary, 394–395
fractional-reserve banking, 466–467
money multiplier, 467–469
100-percent-reserve banking, 466–467
runs on, 473
Barro, David, 59
Bar graph, 40–41
Barriers to entry, sources of, 288
Barter, 457–458
Base year, 333
CPI and, 347
Basket of goods, 346, 349
Baum, L. Frank, 499
Bear Stearns, 471
Bechara, Antoine, 425
Behavioral economics, 424–425
Belchick, Bill, 33
Bernanke, Ben, 464, 470–471, 554
Bernard, Andrew, 340
Bernasek, Anna, 98
Birdsall, Nancy, 380
Blacks
labor-force participation rate, 432
Bloomberg, Michael, 234–235
Bolivia, underground economy in, 336, 337
Bonds, 392, 392–393
Botswana
elephant poaching, 237
Brainard, Lael, 442
Brain drain, 381
Brazil
GDP of, 367
living standard in, 339
Bremer, L. Paul, III, 461
Broadway shows and price discrimination, 306–307
Brodsky, Richard, 235
Broken window fallacy, 15
Bryan, William Jennings, 499
Budget, balanced. See Balanced budget
Budget deficit, 398, 405
crowding out, 406
effect on market for loanable funds, 405–407
of United States, 407–409
Budget surplus, 398, 405
effect on market for loanable funds, 407
Buffett, Warren, 370
Bureau of Labor Statistics (BLS), 430
computing CPI, 346–348
Costs (continued)  
production and, 247–249  
shoeleather, 494–495  
in short- and long-run, 256–257  
social, 206–207  
sunk, 272, 272–273  
total, 244, 252, 259  
transaction, 219  
variable, 251, 252, 259  
Council of Economic Advisers, 31, 32–33  
CPI. See Consumer price index (CPI)  
Credit cards, 463  
Cross-price elasticity of demand, 99, 99  
Crowding out, 406, 559, 559–560  
Cummings, Vincent, 310–311  
Cummins, David, 425  
Deadweight loss, 160–172, 163  
changes in welfare and, 161–163  
determinants of, 164–167  
elasticity and, 164–166  
gains from trade, 163–164  
labor tax, 166–167  
of monopoly, 299–301  
of tariff, 184–185  
tax effects on market participants and, 161–163  
tax revenue and, 167–171  
DeBeers, 289  
Debit cards, 463  
Debt, government. See Government debt  
Debt finance, 393  
Default, on bonds, 393  
Deficits  
budget. See Budget deficit  
Deflation, 357, 479  
Demand, 67–72. See also Equilibrium  
aggregate. See Aggregate demand;  
Model of aggregate demand and  
aggregate supply  
changes in, 79  
decrease in, 69, 70  
elastic, 90, 93  
elasticity of. See Demand elasticity  
equilibrium of supply and demand, 77–82  
excess, 78  
extpectations and, 71  
income changes, 70  
increase in, 69, 70  
individual, 68–69  
inelastic, 90, 93  
for labor. See Labor demand  
law of, 67  
of loanable funds, 400–401  
market, 68–69  
umber of buyers and, 71  
perfectly elastic, 93, 94  
perfectly inelastic, 93, 94  
price of related goods and, 70  
reducing smoking, 71–72  
relationship between price and quan- 
tity demanded, 67–68  
shifts in long and short run, in com- 
petitive markets, 279–282  
taste and, 70  
Demand curve, 42–44, 67–71, 68  
aggregate. See Aggregate-demand  
curve  
measuring consumer surplus using,  
139–140  
price elasticity of demand and, 92–94  
shifts in, 69–71, 279–282  
Demand deposits, 462  
Demand elasticity, 90–99  
aplications of, 102–108  
cross-price elasticity of demand, 99  
deathless loss and, 164–166  
income, 97, 99  
price, 90, 90–99  
Demand schedule, 67  
Denmark  
unemployment benefits, 443  
Depreciation, 329  
Depressions, 507. See also Economic  
fluctuations; Great Depression  
DeSouza, Andrew, 518  
Developed countries  
free trade and wages, 192–193  
Developing countries  
free trade and wages, 192–193  
Diminishing marginal product; 249  
Diminishing returns, 377  
economic growth and, 377–378  
Dinar, 460–461  
Discount coupons, 306  
Discounting, 415  
Discount rate, 471, 471–472  
Discouraged workers, 436  
Discrimination  
price. See Price discrimination  
Diseconomies of scale, 257  
Disposable personal income, 329  
Diversification, 419, 419–420  
Dividends, 395, 422  
Dot-com bubble, end of, 535  
Double coincidence of wants, 458  
Dow Jones Industrial Average, 394  
Drug interdiction, supply, demand and  
elasticity, 106–108  
Dugger, Celia W., 380  
Economic fluctuations, 507–514, 529–539.  
See also Depressions; Great Depression;  
Recession  
assumptions of classical economics,  
510, 512  
irregular and unpredictable nature  
of, 508  
rising unemployment with falling  
output, 510  
shifts in aggregate demand as cause  
of, 530–535  
shifts in aggregate supply as cause of,  
535–539  
short-run, 510–514  
simultaneous fluctuation of macroeco- 
nomic quantities, 508–510  
steps for analyzing, 531  
trash as indicator of, 511  
in United States, 508–509  
Economic growth  
around world, 366–369  
diminishing returns and catch-up  
effect, 377–378  
education and, 379–381  
foreign investment, 378–379  
free trade and, 383  
health and nutrition, 381–382  
Industrial Revolution, 386–387  
long-run, and model of aggregate  
demand and aggregate supply,  
523–524  
natural resources as limit to, 374–376  
population growth and, 384–388  
production possibilities frontier and, 28  
productivity and, 370–371  
property rights and political stability,  
382–383  
public policy and, 374–387  
research and development, 384  
savings and investment, 376  
Economic models, 23–28. See also specific  
models  
Economic profit, 246  
market entry and, 277–278  
Economic Report of the President, 31  
Economics, 4  
environmental, 37  
reasons for studying, 14–15  
Economic welfare  
consumer surplus as measure of,  
141–142  
externalities and, 205–209  
gains and losses from international  
trade, 180–183  
GDP and, 335–338  
monopolies and, 298–302  
tax effect on, 161–163  
Economies  
centrally planned, 9  
market, 9–10  
Economies of scale, 257  
international trade and, 188  
monopoly and, 290–291  
Economist  
vs. accountants and opportunity costs,  
245  
disagreement among, 34–36  
green, 37  
as policy adviser, 30–33  
as scientist, 22–30  
Education. See also Human capital  
cost of college, 5–6  
economic growth and, 379–381  
as externality, 379–380  
itangible capital, 375  
as positive externality, 207  
Edwards, John, 189  
Efficiency, 5, 147, 147–154  
government intervention and, 11–12  
market equilibrium and, 148–152  
market failure and, 152, 154  
}
INDEX

Gains from trade (continued)
of exporting countries, 180–181
of importing countries, 181–183
production possibilities frontier, 50–54
specialization, 55–56
Gallegos, Raul, 122–123
Game-playing factories, 59
Gasoline. See Oil industry
Gas tax
as corrective tax, 211–212
road congestion and, 211, 234–236
GDP. See Gross domestic product (GDP)
GDP deflator, 333, 333–334
computing inflation rate with, 352–353
vs. consumer price index, 352–353
inflation rate and, 334
General Agreement on Tariffs and Trade (GATT), 192–193
General Theory of Employment, Interest, and Money, The (Keynes), 538, 547, 562
George, Henry, 169
Germany
GDP of, 367
hyperinflation in, 488–489
inflation in, 13
labor tax, 170
living standard in, 339, 365
outsourcing, 189
unemployment benefits, 443
Gold standard, 459
Gone with the Wind, 355
Goods, 70
in basket of, for CPI, 349
complements, 70
excludability of, 226
final, 326
inferior, 70
intermediate, 327
markets for, 24–25
price of related, and demand, 70
private, 226
public, 226, 226–230
rivalry of, 226
substitutes, 70, 90
Goolsbee, Austan, 9
Goulden, Lawrence, 37
Government
benefits of, 10–12
monopolies created by, 289–290
Government debt, 405
crowding out, 406
of United States, 407–409
Government policy. See also Public policy for externalities, 209–215
price ceilings, 114–118
price floors, 118–123
Government purchases, 330
aggregate demand and, 555
aggregate-demand curve shifts from, 519
in GDP, 330
Graphs, 40–48
cause and effect, 46–48
curves and, 42–44
measuring profit by, 275–276
of single variable, 40–41
slope of, 44–46
of two variables, 41–42
Great Britain
unemployment benefits, 443
unilateral approach to free trade, 192
Great Britain, health and nutrition in,
381–382
Great Depression, 345
aggregate demand shift and, 533–534
bank runs during, 473
Greece
unemployment benefits, 443
Green economists, 37
Greenpeace, 37
Greenspan, Alan, 426, 472, 554
Gross domestic product (GDP), 324–340,
326
components of, 328–331
consumption in, 329
economic growth and. See Economic growth
as economy’s income and expenditures, 324–325
exclusions from, 326, 337–338
GDP deflator, 333–334
government purchases in, 330
international differences in, 338–339
investment in, 329–330
measurement of, 326–328
as measure of economic well-being, 335–338
net exports in, 330–331
nominal, 332–333
real. See Real GDP
seasonal adjustment, 327–328
underground economy and, 336–337
of United States, 331
winners at Olympic Games and, 340
Gross national product (GNP), 329
“Guns and butter” tradeoff, 4
Hager, George, 511
Hastert, Dennis, 189
Health
economic growth and, 381–382
efficiency wages and, 448
Health insurance, 418
Hedonics, 350–351
Helium market, 83
Holmes, Oliver Wendell, Jr., 159
Hong Kong
free trade and wages, 193
Hoover, Herbert, 345, 354
Households
in circular-flow diagram, 24–25,
324–325
decisions faced by, 3
Housing
in basket of goods for CPI, 349
rent control, 117–118, 122–123
Human capital, 372
economic growth and, 379–381
education as, 380
health and nutrition as, 381–382
per worker, and productivity, 372
Hume, David, 480, 485, 512
Hungary, hyperinflation in, 488–489
Hyperinflation, 480
in Germany, 488–489
money and prices during, 488–489
Implicit cost, 245, 259
cost of capital as, 245–246
Import quota, 36
compared to tariff, 185
Imports. See also International trade
Gains and losses of importing countries, 181–183
Incentives, 7, 7–8
incentive pay, 9
Income
comparing dollar figures from different times, 354
incentive pay, 9
measures of, 329. See also Gross domestic product (GDP); Real GDP
personal, 329
shifts in demand and, 70
Income elasticity of demand, 97, 97, 99
Income tax. See Tax entries
Indexation, 354, 354–355
Index funds, 396
random walk, 423–425
India
GDP of, 367
living standard in, 339, 365
Individual demand, 68–69
vs. market demand, 68–69
Individual income tax. See Tax entries
Individual supply, vs. market supply, 73–74
Indonesia
GDP of, 367
living standard in, 339, 365
Industrial organization, 243–244
Industrial policy, 209
Industrial Revolution, 386–387
Industries. See specific industries
Inefficiency
externalities and, 204–209
Inelastic demand, 90, 93
Inelastic supply, 100, 101
Infant-industry argument for trade restriction, 190–191
Inferior good, 70
income elasticity of demand, 97, 99
Inflation, 12, 13–14, 479–500. See also Hyperinflation
arbitrary redistribution of wealth due to, 488
classical dichotomy and, 485–486
classical theory of, 480–492
comparing dollar figures from different times, 354
confusion and inconvenience, 497–498
costs of, 492–500
deflation, 357, 479
historical perspective of, 479–480
indexation, 354–355
inflation-induced tax distortions, 496–497
inflation tax, 489–490
menu costs, 495
model of aggregate demand and aggregate supply, 523–524
monetary injection and, 482–484
monetary neutrality and, 486
money supply, money demand and monetary equilibrium, 481–482
Intuit, 288, 308
Inventory, GDP and, 330
Investment, 329
aggregate-demand curve shifts from, 515, 517–518
economic growth and, 376
earned income, 282–283
income growth and, 376
interest and, 252–253
interest-rate effect, 546, 549–550
Japan, economic growth in, 367
GDP of, 367
inflation in, 480
labor tax, 170
living standard in, 365
marginal cost, 6
labor market
minimum wage and, 119–121
Labor tax
deadweight loss of, 166–167
in various countries, 170
LaFleur, Tim, 350–351
Laitzare-faire policy, 150
Landsburg, Steven, 187, 404
Land tax, 169
Law of demand, 67
Law of supply, 73
Law of supply and demand, 78
Lender of last resort, 464
Leslie, Phillip, 306–307
Lighthouses, as public good, 230
Liquidity, 459
time of liquidity preference, 547, 547–553
Livingstone, Ken, 235
Loanable funds. See Market for loanable funds
Long run
aggregate-supply curve, 520–523
costs in, 256–257
decision to exit or enter market, 274–275
demand shifts in, 279–282
interest rates in, 550
market supply with entry and exit, 277–278
rent control in, 117–118
Losses
deadweight. See Deadweight loss of exporting countries, 180–181
of importing countries, 181–183
Lucas, Robert, 366
Luxuries
income elasticity of demand, 99
price elasticity of demand and, 90
Luxury tax, 130

INDEX 579

price level and value of money, 480–481
purchasing power fall due to, 493–494
relative-price variability and misallocation of resources, 495–496
shoeleather costs, 494–495
tradeoff with unemployment, 14–15
velocity of money and quantity equation, 486–488
Inflation rate, 348
computing with CPI, 348, 352–353
computing with GDP deflator, 352–353
GDP deflator and, 334
real and nominal interest rates and, 355–356
Inflation tax, 489–490, 490
Informational efficiency, 423
Inputs. See Factors of production; specific factors of production
An Inquiry into the Nature and Causes of the Wealth of Nations (Smith), 10, 57, 258
Insurance
health, 418
markets for, 417–419
role of, 417–419, 418–419
unemployment, 441, 441–443
Intangible capital, 374–375
Interest
compounding, 414
Interest rate(s)
discount rate, 471–472
equilibrium, 549
federal funds rate, 473–474
in long run, 550
nominal, 355–356, 356
real, 355–356, 356
in short run, 550
supply and demand for loanable funds, 400–401
target, in Fed policy, 553
theory of liquidity preference, 547–553
in United States, 357
Interest-rate effect, 546, 549–550
aggregate-demand curve and, 515
Intermediate good, 327
Internalizing the externality, 207
International Monetary Fund, 379
International trade, 177–195
benefits of, 186, 188
comparative advantage. See Comparative advantage
determinants of, 178–179
effect of tariff on, 183–185
equilibrium without, 178–179
free trade, 192–194
gains and losses of exporting countries, 180–181
gains and losses of importing countries, 181–183
multilateral approach to free trade, 192–194
outsourcing, 187, 189, 192
restriction on. See Trade restrictions
trade agreements and trade organizations, 192–195
trade policy and, 185–186
of United States, 38–59
world price and, 179
Labor
aggregate-supply curve shifts and, 522
jobs argument for trade restriction, 188–190
taxes on, 166–167, 170
Labor demand
minimum wage and, 120–121
Labor force, 431
Labor-force participation rate, 431, 431–433
for men, 432, 433–435
for women, 432, 433–435
Labor market
minimum wage and, 119–121
Labor tax
deadweight loss of, 166–167
in various countries, 170
LaFleur, Arthur, 169–171
LaFleur curve, 171
LaFleur, Tim, 350–351
Laissez-faire policy, 150
Landsburg, Steven, 187, 404
Land tax, 169
Law of demand, 67
Law of supply, 73
Law of supply and demand, 78
Lender of last resort, 464
Leslie, Phillip, 306–307
Lighthouses, as public good, 230
Liquidity, 459
time of liquidity preference, 547, 547–553
Livingstone, Ken, 235
Loanable funds. See Market for loanable funds
Long run
aggregate-supply curve, 520–523
costs in, 256–257
decision to exit or enter market, 274–275
demand shifts in, 279–282
interest rates in, 550
market supply with entry and exit, 277–278
rent control in, 117–118
Losses
deadweight. See Deadweight loss of exporting countries, 180–181
of importing countries, 181–183
Lucas, Robert, 366
Luxuries
income elasticity of demand, 99
price elasticity of demand and, 90
Luxury tax, 130

M1, 462
M2, 462
Macroeconomics, 28, 28–30, 324
Malawi
elephant poaching, 237
Mali, living standard in, 368–369
Malnutrition, 382
Malthus, Thomas Robert, 384–385, 386–387
Marginal benefit, 6
Marginal changes, 6
Marginal cost, 6, 252, 259
marginal-cost pricing for natural monopoly, 311–312
related to average total cost, 254
rising, 253
Marginal cost curve, 253, 255
firm’s supply decision and, 269–271
Marginal product, 248
diminishing, 249
Marginal propensity to consume, multiplier effect and, 557–558
Marginal revenue, 268
of monopoly, 293–294
Market(s), 66. See also specific markets
competitive. See Competitive markets
definition of, and price elasticity of
demand, 91
for factors of production, 24–25
for goods and services, 24–25
irrationality of, 425–426
Market-based policy, 210–215
Market demand, 68–69
vs. individual demand, 68–69
Market economy, 9
Market efficiency. See Efficiency
Market failure, 11. See also Externalities
efficiency and, 154
Market for loanable funds, 399, 399–407
crowding out, 406
government budget deficits and sur-
pluses effect on, 405–407
investment incentives, 403–405
saving incentives, 402–403
supply and demand, 400–401
Market power, 11, 152, 154
Market risk, 420
Market supply
in competitive markets. See Competitive
markets
vs. individual supply, 73–74
Martin, William McChesney, 563
Mass transit system, public vs. private,
310–311
McCain, John, 187
McKinley, William, 499
McTeer, Robert D., Jr., 14–15
Medicare tax, 166, 170
Medium of exchange, 395, 459
Men, labor-force participation rate, 432,
433–435
Menu costs, 495, 495, 526
Metcalf, Gilbert, 216
Mexico
free trade and wages, 193
GDP of, 367
living standard in, 12, 339, 368–369
multilateral approach to free trade,
192–193
underground economy in, 336
Microeconomics, 28, 28–30, 323, 323–324
Microsoft Corporation, 287–288, 308
Midpoint method, for calculating price
elasticity of demand, 91–92
Miller, Peter, 122–123
Minimum wage
characteristics of workers earning, 445
opponents/advocates of, 121
as price floor, 119–121
teenage employment and, 120–121
Minimum-wage laws, 443–445
Misperceptions theory, aggregate-supply
curve, 527
Model of aggregate demand and aggregate
supply, 513, 513–539
aggregate-demand curve, 514–519
aggregate-supply curve, 519–529
causes of economic fluctuations and,
529–539
long-run growth and inflation,
523–524
monetary neutrality and, 532
Monetary equilibrium, inflation and,
481–482
Monetary neutrality, 486, 532
Monetary policy, 464
effects of monetary injection, 482–484
free-silver debate, 499–500
influence on aggregate demand,
546–555
lag in, 564
during recession of 2001, 535
role of interest-rate targets in Fed
policy, 553
stabilization policy arguments,
562–564
Monetary system
banks and money supply, 466–474
Federal Reserve (Fed), 464–465
meaning of money, 458–463
Money, 457–474, 458. See also Federal
Reserve System (Fed); Monetary policy;
Money supply
commodity, 459, 459–460
creation with fractional-reserve bank-
ing, 466–467
fiat, 460, 460–461
functions of, 458–459
future value, 414
in Iraq, 460–461
kinds of, 459–461
present value, 414–416
quantity theory of, 483
in U.S. economy, 461–463
velocity of, 486, 486–488
Money demand
inflation and, 481–482
theory of liquidity preference, 548–549
Money market, equilibrium in, and theory
of liquidity preference, 549
Money multiplier, 467–469, 468
Money supply, 464
aggregate-demand curve shifts from,
517–518
bank runs and, 473
changes in, and aggregate-demand
curve, 551–553
creation with fractional-reserve bank-
ing, 466–467
discount rate, 471–472
Fed’s tools for monetary control,
469–472
Great Depression and, 533–534
inflation and, 13–14, 481–482
money multiplier, 467–469
open-market operations, 469
problems in controlling, 472
reserve requirements, 470–471
theory of liquidity preference, 547–548
Monopoly, 67, 287–314, 288
antitrust laws, 308–309
barriers to entry and, 288–291
vs. competitive firm, 291–292, 313
deadweight loss, 299–301
economies of scale and, 290–291
efficiency and, 299–301
government-created, 289–290
lack of supply curve for, 296
natural, 290, 290–291
prevalence of, 314
price and, 291–292, 293
price discrimination. See Price
discrimination
profit maximization, 295–296
profit of, 295–297, 301–302
public ownership of, 312
reasons for, 288–291
regulation, 309–312
resources and, 289
revenue of, 292–294
welfare costs of, 298–302
Moral hazard
insurance and, 419
Movie industry, movie tickets as price
discrimination, 305
Mugabe, Robert G., 491
Multiplier effect, 556
aggregate demand and, 556–558
other application of, 558
Municipal bonds, 393
Muskie, Edmund, 214
Mutual funds, 395
as financial intermediary, 395–396
random walks and index funds,
423–425
Nader, Ralph, 7
NAFTA, 192
Namibia
elephant poaching, 237
NASDAQ (National Association of
Securities Dealers Automated Quotation
system), 394
National defense
as public good, 228–229
spending on, 228–229
National Highway Traffic Safety
Administration, 211
National income, 329
National income accounts, 397–399
National Institutes of Health, 229, 384
National Labor Relations Board (NLRB),
447
National saving, 398, 398–399
National Science Foundation, 229, 384
National security argument for trade
restriction, 190
Natural disasters
prices and, 84
Natural monopoly, 227, 290, 290–291,
310–311
Natural rate of output, 521
Natural rate of unemployment, 430, 433,
433
Natural resources, 372
aggregate-supply curve shifts and, 522
as limit to growth, 374–376
monopolies and, 289
per worker, and productivity, 372
population growth and stretching,
384–385
Necessities, price elasticity of demand and,
90
Negative correlation, 42
Negative externality, 204, 205–207
market supply with fixed number of firms, 277
rent control in, 117–118
Shutdown, short-run decision to, 271–273
Sierra Club, 215
Silvia, John, 443
Sinaï, Allen, 511
Singapore
free trade and wages, 193
international trade of, 383
living standard in, 365
Sinn, Hans-Werner, 442
Slemrod, Joel, 518
Slope, 44–46
Smith, Adam, 10, 11, 57, 84, 153, 258
Smoking, reducing, 71–72
Social Security
Social costs, of externalities, 206–207
Stagflation, 536
Standard of living
determinants of, 12–13
GDP and, 339
intangible capital, 374–375
large variation in, 365–369
productivity’s role in, 370–371
Standard & Poor’s 500 Index, 394
Statistical discrepancy, 328
Stein, Charles, 151
Stewart, Jay, 434, 435
Stewart, Wayne, 33
Sticky-price theory, aggregate-supply curve
and, 526
Sticky-wage theory, 529, 536
aggregate-supply curve, 525–526
Stigler, George, 313
Stock index, 394
Stockman, David, 171
Stock market, 393–394
end of dot-com bubble, 535
Fed monitoring, 554–555
reading newspaper stock tables, 395
Stocks, 393
efficient markets hypothesis, 422–423
emotions and investing decisions, 424–425
fundamental analysis, 422
market irrationality, 425–426
random walks and index funds, 423–425
speculative bubble, 426
Store of value, 395, 459
Stossel, John, 84
Strike, 446
Structural unemployment, 438
minimum-wage laws and, 443–444
Substitutes, 70
cross-price elasticity of demand, 99
price elasticity of demand and, 90
Substitution bias, 348–349
Sullivan, John, 518
Sunk cost, 272, 272–273
Supply, 73–76. See also Equilibrium aggregate.
See Model of aggregate demand and aggregate supply changes in, 80–81
decrease in, 74, 75
elastic, 100, 101
elasticity of, 99–102
equilibrium of supply and demand, 77–82
excess, 77–78
expectations and, 76
increase in, 74, 75
individual, 73–74
inelastic, 100, 101
input prices and, 74, 76
law of, 73
of loanable funds, 400–401
marginal-cost curve and firm’s supply decision, 269–271
market, 73–74
number of sellers, 76
perfectly elastic, 100, 101
perfectly inelastic, 100, 101
relationship between price and quantity supplied, 73
technology and, 76
Supply and demand, law of, 78
Supply curve, 73, 73–76
aggregate. See Aggregate-supply curve in competitive market, 276–282
lack of, for monopoly, 296
long-run, upward-sloping, 280–282
measuring producer surplus with, 144–145
price elasticity of supply, 100–102
shifts in, 74–76, 80
shifts in vs. movement along, 79–80
Supply elasticity, 99–102
applications of, 102–108
deadweight loss and, 164–166
Supply schedule, 73
Supply-side economics, 171, 561
Surplus, 77
budget, 398
consumer. See Consumer surplus
price floor, 119, 120
producer. See Producer surplus
total surplus and efficiency, 147–148
Sweden
tax rate in, 171
underground economy in, 336
Switzerland
intangible capital, 375
rule-of-law index, 375
unemployment benefits, 443
Synergies, 308
T-account, 466
Taiwan
free trade and wages, 193
international trade of, 383
living standard in, 365
Tanzania
elephant poaching, 237
Tariffs, 183
compared to import quotas, 185
deadweight loss of, 184–185
effects of, 183–185
Tastes
shifts in demand and, 70
Taxes, 123–130, 159–172
aggregate demand and, 517, 560–561
aggregate supply and, 561
automatic stabilizers, 565
on bonds, 393
on buyers, effect on market outcomes, 125–127
carbon tax, 216–217
corrective, 210
cuts under G. W. Bush, 535, 560–561, 563
cuts under Kennedy, 563
cuts under Reagan, 171
deadweight loss of, 161–163, 167–172
fiscal policy and, 560–561
gas tax, 211–212, 234–236
health insurance and, 418
incidence, 124, 128–130
inflation-induced distortions of, 496–497
inflation tax, 489–490
investment incentives, 403–405
on labor, 166–167, 170
Laffer curve and supply-side economics, 169–171
on land, 169
luxury, 130
payroll, 127
revenue from, and deadweight loss, 167–171
saving incentives and, 402–403
on sellers, effect on market outcome, 124–125
tariff, 184–185
tax rebate stimulus package, 518
Tax incidence, 124
elasticity and, 128–130
Technological knowledge, 372
Technology
aggregate-supply curve shifts and changes in, 322–323
natural resources limitations and changes in, 375–376
population growth to promote technological progress, 387–388
research and development and economic growth, 384
supply curve shifts and, 76
technological knowledge and productivity, 372–374
Technology (continued)
Technology spillovers, 208–209
Teenage employment, minimum wage and, 120–121
Temporary Assistance for Needy Families (TANF), 229
Term, of bonds, 393
Term Auction Facility (TAF), 471
Term Securities Lending Facility (TSLF), 471
Textile industry, 177
Thailand
underground economy in, 336, 337
Theory, 22–23
Theory of liquidity preference, 547, 547–553
Ticket scalping, 151
Tierney, John, 310–311
Ticket scalping, 151
Theory, 22–23
Theory of liquidity preference, 547, 547–553
Ticket scalping, 151
Tierney, John, 310–311
Time horizon. See also Long run; Short run
price elasticity of demand and, 91
price elasticity of supply and, 100
Time-series graph, 41
Tobin, James, 563
Toll roads, 235
Total cost, 244, 259
average, 252, 253, 254
Total-cost curve, production function and, 248–249
Total revenue, 94, 244
along linear demand curve, 95–97
price elasticity of demand and, 94–95
Total surplus, efficiency and, 147–148
 Tradable permits, 37
Tradable pollution permits, 212–214, 217
Trade
benefits of, 8
comparative advantage and, 55–56
economic growth and, 383
gains from. See also Gains from trade international. See International trade; Trade restrictions
specialization and, 52–54
Tradeoffs, 4–5
policy decisions and, 31
production possibilities frontier and, 27
between risk and return, 420–421
Trade restrictions
effects of tariff, 183–185
import quotas, 185
infant-industry argument, 190–191
jobs argument for, 188–190
national security argument, 190
protection-as-a-bargaining-chip argument, 191
tariffs, 36, 183–185
unfair-competition argument, 191
Traffic
congestion and common resource, 234–236
congestion pricing, 234–235
Tragedy of the Commons, 232, 232–233
Transaction costs, 219
Transfer payment, 330
Transportation
in basket of goods for CPI, 349
incentive pay for bus drivers, 9
Transport Workers Union, 311
Truman, Harry, 31
Turnover, efficiency wages and, 448–449
Uganda
elephant poaching, 237
Underground economy, 167, 336–337
Unemployment, 429–450
cyclical, 430, 433
duration of, 437–438
frictional, 438, 439–440, 441
Great Depression, 534
job search and, 439–442
measurement of, 430–435
minimum-wage laws and, 121, 443–445
natural rate of, 433
rate. See Unemployment rate reasons for, 438
rising, with falling output, 510
structural, 438, 443–444
Unemployment benefits
in various countries, 442–443
Unemployment insurance, 441
as automatic stabilizer, 565
in various countries, 442–443
Unemployment rate, 431
aggregate-supply curve shifts and, 522
alternative measures of, 437
computation of, 431
discourage workers, 436
natural rate of, 430, 433
in United States, since 1960, 433, 509
Unfair-competition argument for trade restriction, 191
Unions, 445, 445–447
economics of, 446–447
reasons for/against belonging to, 447
United Kingdom
GDP of, 367
labor tax, 170
living standard in, 367, 368
underground economy in, 336
United States
cap-and-trade system, 217
carbon tax, 217
currency in, 463
economic fluctuations, 508–509
economic growth in, 366–367
financial institutions in, 352–396
GDP of, 331, 367
government debt of, 407–409
intangible capital, 375
interest rate in, 357
international trade and, 58–59
investment spending, 509
labor tax, 170
living standard in, 12, 339, 365
money in, 461–463
multilateral approach to free trade, 192–193
NAFTA, 192
real GDP, 334–335, 509
recessions in, 508
rule-of-law index, 375
underground economy in, 336–337
unemployment benefits, 442–443
unemployment rate in, 433, 509
Unit of account, 459
Unsafe at Any Speed (Nader), 7
U.S. Postal Service, 312
Utility, 417
Value. See also Asset valuation
store of, 459
Values, differences among economists in, 34–35
Variable costs, 251, 259
average, 252
Variables
graphs of single, 40–41
graphs of two, 41–42
nominal, 485
omitted, causality and, 46–47
real, 485
Vedantam, Shankar, 32–33
Velocity of money, 486, 486–488
Venezuela
inflation in, 480
price controls, 122–123
Wade, Nicholas, 386–387
Wage-price spiral, 537
Wages
efficiency, 448, 448–450
free trade and, 192–193
minimum, 119–121
minimum-wage laws, 443–445
sticky, 525–526
Wage subsidies, 123
Wagner Act, 446–447
Water, clean, as common resource, 233–234
Wealth effect, 546
aggregate-demand curve and, 515
Wealth of Nations, The (Smith), 11
The Wealth of Nations (Smith), 258
Welfare economics, 137, 205
White, Matthew W., 98
Wildlife, as common resource, 236
Will, George F., 189
Willingness to pay, 138, 138–139
Willingness to sell, producer surplus and, 143–144
Windows operating system, 287–288
Wines, Michael, 491
Wizard of Oz (Baum), 498–500
Women
labor-force participation rate, 432, 433–435
World Bank, 374–375, 379
WorldCom, 535
World price, 179
World Trade Organization (WTO), 193
World War II, aggregate demand shift and, 533–534
x-coordinate, 41
y-coordinate, 41
Zimbabwe
elephant poaching, 237
inflation in, 480, 491
underground economy in, 336, 337