Microeconomics

Microeconomics

IVAN VELASCO, SANTA ANA COLLEGE

Contents

Part XII. Module: Monopolistically Competitive Industry

Part XIII. Module: Oligopoly

Course Learning Outcomes

Microeconomics Learning Outcomes

The Learning Outcomes covered by these Microeconomics course materials are:

Module (Chapter)	Learning Outcomes
Economic Thinking: Prepare for success in studying economics	• Explain what economics is and explain why it is important • Explain how economists use economic models • Use mathematics in common economic applications • Use graphs in common economic applications
Choice in a World of Scarcity: Use economic thinking to explain choice in a world of scarcity	• Explain the cost of choices and trade-offs • Illustrate society's trade-offs by using a production possibilities frontier (or curve) • Explain the assumption of rationality by individuals and firms • Define marginal analysis • Differentiate between positive and normative statements
Supply and Demand: Analyze how buyers and sellers interact in a free and competitive market to determine prices and quantities of goods	• Describe and differentiate between major economic systems • Explain the determinants of demand • Explain the determinants of supply • Explain and graphically illustrate market equilibrium, surplus and shortage
Elasticity: Measure how changes in price and income affect the behavior of buyers and sellers	• Explain the concept of elasticity • Explain the price elasticity of demand and price elasticity of supply, and compute both using the midpoint method • Explain and calculate other elasticities using common economic variables • Explain the relationship between a firm's price elasticity of demand and total revenue

Government Action: Evaluate the consequences of government policies in markets	• Analyze the consequences of the government setting a binding price ceiling • Analyze the consequences of the government setting a binding price floor • Explain how the price elasticities of demand and supply affect the incidence of a sales tax • Define progressive, proportional, and regressive taxes
Surplus: Use the concept of producer, consumer surplus, and total surplus to explain the outcomes of markets for individuals, firms, and society	• Define and calculate consumer, producer and total surplus; graphically illustrate consumer, producer and total surplus • Use the concepts of consumer, producer and total surplus to explain why markets typically lead to efficient outcomes
Utility: Explain how consumer behavior shapes the demand curve with respect to utility and loss	• Define the concept of utility and satisfaction • Differentiate between marginal utility and total utility • Describe and calculate the concept of marginal utility • Explain how consumers maximize total utility within a given income using the Utility Maximizing Rule • Explain how consumer's utility changes when income or prices change • Describe the behavioral economics approach to understanding decision making

Production: Analyze the relationship between inputs used in production and the resulting outputs and costs	• Define the term “production” and explain what a production function is; define the term “production inputs,” and differentiate between labor, land, capital, entrepreneurship, technology • Define and differentiate between marginal, average, and total product; compute and graph marginal, average, and total product; explain diminishing marginal product and diminishing marginal returns • Differentiate between Explicit and Implicit Costs, Accounting and Economic Profit • Identify sunk costs • Define and differentiate between marginal, average, and total cost; compute and graph marginal, average, and total cost; differentiate between variable and fixed costs • Differentiate between short-run and long-run costs; interpret the relationship between short-run and long-run costs • Define and explain long-run costs, economies of scale, diseconomies of scale, and constant returns to scale

Perfect Competition: Analyze a firm's profit maximizing decisions under conditions of perfect competition	• Define the characteristics of Perfect Competition • Understand the difference between the firm and the industry; explain and illustrate the differences between the demand curve for a perfectly competitive firm and that for a perfectly competitive industry • Calculate and graph the firm's fixed, variable, average, marginal and total costs; calculate and graph the firm's average, marginal and total revenues; determine the profit maximizing output level and price using graphs and demand schedules; is able to calculate and graphically illustrate where marginal revenue equals marginal costs • Measure variable and total costs as the area under the average variable and average total cost curves; measure total revenues as the area under the average revenue curves; calculate and graphically illustrate profit and losses for a perfectly competitive firm • Determine the break-even, and the shutdown points of production for a perfectly competitive firm • Explain the difference between short-run and long-run equilibrium; explain the concept of "zero economic profit" • Understand why perfectly competitive markets are efficient
Monopoly: Analyze a firm's profit maximizing strategies under conditions of a monopoly	• Define the characteristics of a monopoly • Define and explain the sources of barriers to entry • Calculate and graph a monopoly's fixed, variable, average, marginal and total costs; measure variable and total costs as the area under the average variable and average total cost curves; calculate and graph the firm's average, marginal and total revenues; measure total revenues as the area under the average revenue curves; determine the profit maximizing output level and price; is able to calculate and graphically illustrate where marginal revenue equals marginal costs; calculate and graphically illustrate profit and losses for a monopolist • Explain why a monopoly is inefficient using deadweight loss; differentiate between a single price monopolist and a price discriminating monopolist • Analyze different strategies to control monopolies, including natural monopolies

Oligopoly: Analyze a firm's profit maximizing strategies under conditions of oligopoly	• Define characteristics of oligopolies • Explain why collusion can occur in oligopolistic industries • Explain the role of game theory in understanding the behavior of oligopolies • Explain why oligopolies are inefficient
Monopolistically Competitive Industries: Analyze a firm's profit maximizing strategies under conditions of monopolistic competition	• Define the characteristics of a monopolistically competitive industry; understand the difference between the firm and the industry • Calculate and graph the firm's fixed, variable, average, marginal and total costs; measure variable and total costs as the area under the average variable and average total cost curves; calculate and graph the firm's average, marginal and total revenues; measure total revenues as the area under the average revenue curves; determine the profit maximizing output level and price; is able to calculate and graphically illustrate where marginal revenue equals marginal costs; calculate and graphically illustrate profit and losses for a monopolistically competitive firm • Explain the difference between short run and long run equilibrium in a monopolistically competitive industry • Understand how product differentiation works in monopolistically competitive industries and how firms use advertising to differentiate their products, understanding impact on elasticity • Understand why monopolistically competitive markets are inefficient (including deadweight loss)
Public Goods: Compare public goods and private goods and understand the role for them in the economy	• Contrast between public and private goods • Explain the concept of free riders • Define and give examples of positive and negative externalities • Analyze the efficacy of government policies to lessen negative externalities and analyze how the government promotes positive externalities • Analyze the impact of market-based solutions to negative externalities

Globalization, Trade, and Finance: Analyze the benefits and costs of international trade	• Define and calculate comparative and absolute advantage • Explain how a nation's workers and consumers are affected by impact of international trade • Understand the way government regulations (e.g. tariffs, quotas and non-tariff barriers) affect business, consumers and workers in the economy • Differentiate between alternative international trade regimes and how they impact global trade • Define currency exchange rates and explain how they influence trade balances • Explain how the balance of trade (surplus or deficit) affects the domestic economy, and how the domestic economy affects the balance of trade • Connect globalization, international trade, and international finance
Income Distribution: Assess how resource markets/factors of production affect society's distribution of income	• Describe the incomes earned by the factors of production (land, labor, capital, entrepreneurship) wages, interest, rents, and profit • Analyze how perfect/imperfect competition between buyers and sellers of factors can impact wages, interest, and rents • Use the Lorenz Curve to analyze the distribution of income and wealth

PART I

FACULTY RESOURCES

1. Question Banks

Note: *It is your responsibility to handle question banks and answer keys* ***securely*** *and appropriately to prevent them from becoming widely available and searchable via the Internet.*

The quizzes included in this course are organized by module and are aligned to specific learning outcomes. They are available as a separate file that can be downloaded and imported into the assessment tool in your learning management system (Canvas, Blackboard, etc.).

There are two different quiz download files available. They contain the same questions, organized in a different manner.

- Module quizzes (15 quizzes, with a total of 693 questions)
- Granular quizzes (74 quizzes, with the same 980 questions)

Next, follow the instructions for importing quiz files here: Candela Quiz Imports

2. Additional Resources

OpenStax Additional Resources

OpenStax College, the primary source for openly-licensed materials in this course, requests that instructors register and log in to request access to available instructor resources.

Available instructor resources may include items like:

- Getting Started Guide
- Instructor Solutions Manual
- Lecture Slides
- Sample Syllabus Language
- Test Banks

Supplemental Materials

Khan Academy

3. I Need Help

Need more information about this course? Have questions about faculty resources? Can't find what you're looking for? Experiencing technical difficulties?

We're here to help! Contact oer@achievingthedream.org for support.

PART II
MODULE: ECONOMIC THINKING

4. Why It Matters: Economic Thinking

Why explain the basic premises and tools of economic thinking?

Many students find the prospect of taking an economics course daunting (or maybe just dull). At the heart of this worry is perhaps just a misperception of what economics is about. It's not rocket science, it's not a collection of boring facts, and it's not the study of money or the stock market. Economics is really just a set of interesting questions organized around a simple fact: There aren't enough resources (money, land, time, etc.) to go around or satisfy all our needs and desires. Economists call this condition *scarcity*. It affects individuals, nations, and the entire human species—no one ever has enough of the things they want. On some level, everyone has to grapple with scarcity, and economists are interested in understanding how people do that.

If you understand how people behave in the face of scarcity—and learn to think like an economist—economics can be an amazingly powerful tool. You can predict the behavior of individual economic agents, such as consumers or businesses—what economists call the micro level. You can predict the behavior of an economy (or economies) as a whole—what economists call the macro level. You can have a better understanding of the choices—and consequences—in your own life.

Consider the following example:

Imagine that you're about to catch a flight to Italy. You've saved and saved to pay for this trip, and you're thrilled to finally be going. You're on top of the world, until . . .

You get to the airport and have to go through security. The line is terrible. What choice do you have? You can't board the plane without passing security. As you wait, you notice a different aisle for "special" passengers who fly more often. They aren't waiting at all. In fact, if more than three of them collect in the line and have to wait for more than a minute, they act very annoyed—shifting their weight, rolling their eyes, checking their phones, and so on. Oh, brother! You've been waiting so much longer! How is that fair?

Finally you make it through security and reach your gate. Sadly, you are in zone 5, which boards last. You have to struggle down the aisle—past rows of seats with more leg room—to a center seat. Worse, those who boarded before you have filled all the overhead bins. A flight attendant seems irritated that you have a large carry-on bag that won't fit under your seat. He takes your bag off the plane and tells you to pick it up at baggage claim after the flight.

You tuck in your elbows and squeeze into you seat thinking, This isn't fair.

Not sure what all this has to do with scarcity? You need to study economics!

Learning Outcomes

- Explain what economics is and why it is important
- Explain how economists use economic models
- Use mathematics in common economic applications
- Use graphs in common economic applications

5. Outcome: What Is Economics?

What you'll learn to do: explain what economics is and explain why it is important

In order to understand economics it's important to master a set of key definitions and understand how they interconnect. These concepts will be used many times throughout the course. At the most basic level:

- Scarcity means that there are never enough resources to satisfy all human wants
- Economics is the study of the trade-offs and choices that we make, given the fact of scarcity
- Opportunity cost is what we give up when we choose one thing over another

We will spend more time with these definitions, and understand how they're used in the context of this discipline.

The specific things you'll learn in this section include the following:

- Define scarcity and explain its economic impact
- Define opportunity cost
- Define productive resources
- Explain why trade and markets exist
- Distinguish between macroeconomics and microeconomics

Learning Activities

The learning activities for this section include the following:

- Video: Scarcity and Choice
- Reading: Understanding Economics and Scarcity
- Video: Resources
- Reading: The Concept of Opportunity Cost
- Video: Opportunity Cost
- Reading: Labor, Markets, and Trade
- Reading: Microeconomics and Macroeconomics

6. Video: Scarcity and Choice

Throughout this course you'll encounter a series of short videos that explain complex economic concepts in very simple terms. Take the time to watch them! They'll help you master the basics before heading to the readings (which tend to cover the same information in more depth).

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=25#oembed-1

As you watch the video, consider the following key points:

1. Economics is the study of how humans make choices under conditions of scarcity.
2. Scarcity exists when human wants for goods and services exceed the available supply.
3. People make decisions in their own self-interest, weighing benefits and costs.

7. Reading: Understanding Economics and Scarcity

Scarcity

The resources that we value—time, money, labor, tools, land, and raw materials—exist in limited supply. There are simply never enough resources to meet all our needs and desires. This condition is known as scarcity.

At any moment in time, there is a finite amount of resources available. Even when the number of resources is very large, it's limited. For example, according to the U.S. Bureau of Labor Statistics, in 2016, the labor force in the United States contained more than 158 million workers—that's a lot, but it's not infinite. Similarly, the total area of the United States is 3,794,101 square miles—an impressive amount of acreage, but not endless. Because these resources are limited, so are the numbers of goods and services we can produce with them. Combine this with the fact that human wants seem to be virtually infinite, and you can see why scarcity is a problem.

Economics

When faced with limited resources, we have to make choices. Again, economics is the study of how humans make choices under conditions of scarcity. These decisions can be made by individuals, families, businesses, or societies.

Let's consider a few decisions that we make based on limited resources. Take the following:

1. What classes are you taking this term?

Are you the lucky student who is taking every class you wanted with your first-choice professor during the perfect time and at the ideal location? The odds are that you have probably had to make trade-offs on account of scarcity. There is a limited number of time slots each day for classes and only so many faculty available to teach them. Every faculty member can't be assigned to every time slot. Only one class can be assigned to each classroom at a given time. This means that each student has to make trade-offs between the time slot, the instructor, and the class location.

2. Where do you live?

Think for a moment, if you had all the money in the world, where would you live? It's probably not where you're living today. You have probably made a housing decision based on scarcity. What location did you pick? Given limited time, you may have chosen to live close to work or school. Given the demand for housing, some locations are more expensive than others, though, and you may have chosen to spend more money for a convenient location or to spend less money for a place that leaves you spending more time on transportation. There is a limited amount of housing in any location, so you are forced to choose from what's available at any time. Housing decisions always have to take into account what someone can afford. Individuals making decisions about where to live must deal with limitations of financial resources, available housing options, time, and often other restrictions created by builders, landlords, city planners, and government regulations.

The Problem of Scarcity

Every society, at every level, must make choices about how to use its resources. Families must decide whether to spend their money on a new car or a fancy vacation. Towns must choose whether to put more of the budget into police and fire protection or into the school system. Nations must decide whether to devote more funds to national defense or to protecting the environment. In most cases, there just isn't enough money in the budget to do everything.

Economics helps us understand the decisions that individuals, families, businesses, or societies make, given the fact that there are never enough resources to address all needs and desires.

8. Video: Resources

https://youtu.be/0PgP0dXAGAE

Below are the key points from the video:

There are four **productive resources** (resources have to be able to produce something), also called **factors of production**:

- **Land:** any natural resource, including actual land, but also trees, plants, livestock, wind, sun, water, etc.
- **Economic capital:** anything that's manufactured in order to be used in the production of goods and services. Note the distinction between financial capital (which is not productive) and economic capital (which is). While money isn't directly productive, the tools and machinery that it buys can be.
- **Labor:** any human service—physical or intellectual. Also referred to as *human capital*.
- **Entrepreneurship**: the ability of someone (an entrepreneur) to recognize a profit opportunity, organize the other factors of production, and accept risk.

9. Reading: The Concept of Opportunity Cost

The Idea of Opportunity Cost

Since resources are limited, every time you make a choice about how to use them, you are also choosing to forego other options. Economists use the term **opportunity cost** to indicate what must be given up to obtain something that's desired. A fundamental principle of economics is that every choice has an opportunity cost. If you sleep through your economics class (not recommended, by the way), the opportunity cost is the learning you miss. If you spend your income on video games, you cannot spend it on movies. If you choose to marry one person, you give up the opportunity to marry anyone else. In short, opportunity cost is all around us.

The idea behind opportunity cost is that the cost of one item is the lost opportunity to do or consume something else; in short, opportunity cost is the value of the next best alternative.

Since people must choose, they inevitably face trade-offs in which they have to give up things they desire to get other things they desire more.

Opportunity Cost and Individual Decisions

In some cases, recognizing the opportunity cost can alter personal behavior. Imagine, for example, that you spend $8 on lunch every day at work. You may know perfectly well that bringing a lunch from home would cost only $3 a day, so the opportunity cost of buying lunch at the restaurant is $5 each day (that is, the $8 that buying lunch costs minus the $3 your lunch from home would cost). Five dollars each day does not seem to be that much. However, if you project what that adds up to in a year—250 workdays a year × $5 per day equals $1,250—it's the cost, perhaps, of a decent vacation. If the opportunity cost were described as "a nice vacation" instead of "$5 a day," you might make different choices.

Opportunity Cost and Societal Decisions

Opportunity cost also comes into play with societal decisions. Universal health care would be nice, but the opportunity cost of such a decision would be less housing, environmental protection, or national defense. These trade-offs also arise with government policies. For example, after the terrorist plane hijackings on September 11, 2001, many proposals, such as the following, were made to improve air travel safety:

- The federal government could provide armed "sky marshals" who would travel inconspicuously with the rest of the passengers. The cost of having a sky marshal on every flight would be roughly $3 billion per year.
- Retrofitting all U.S. planes with reinforced cockpit doors to make it harder for terrorists to take over the plane would have a price tag of $450 million.
- Buying more sophisticated security equipment for airports, like three-dimensional baggage scanners and cameras linked to face-recognition software, would cost another $2 billion.

Lost time can be a significant component of opportunity cost.

However, the single biggest cost of greater airline security doesn't involve money. It's the opportunity cost of additional waiting time at the airport. According to the United States Department of Transportation, more than 800 million passengers took plane trips in the United States in 2012. Since the 9/11 hijackings, security screening has become more intensive, and consequently, the procedure takes longer than in the past. Say that, on average, each air passenger spends an extra 30 minutes in the airport per trip. Economists commonly place a value on time to convert an opportunity cost in time into a monetary figure. Because many air travelers are relatively highly paid businesspeople, conservative estimates set the average "price of time" for air travelers at $20 per hour. Accordingly, the opportunity cost of delays in airports could be as much as 800 million (passengers) × 0.5 hours × $20/hour—or, $8 billion per year. Clearly, the opportunity costs of waiting time can be just as substantial as costs involving direct spending.

Opportunity Cost Video

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=28#oembed-1

10. Reading: Labor, Markets, and Trade

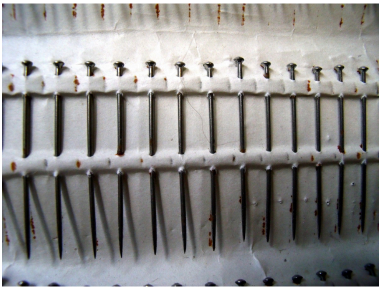
Assembly Line

The Division and Specialization of Labor

We have learned that there aren't enough resources to fulfill all of our wants and this reality forces us to make choices that have opportunity costs. How do we get the most we can from the resources we have? Over time, markets and trade have come into existence and have become highly efficient mechanisms for optimizing our use of resources and bringing us the most and best combination of goods and services.

Think back to pioneer days, when the average person knew how to do so much more on his or her own than someone today—everything from shoeing a horse to growing, hunting, and preserving food to building a house and repairing equipment. Most of us don't know how to do all—or any—of those things. It's not because we're not capable of learning them. It's because we don't have to. The reason for this is something called the "division and specialization of labor," a production innovation first put forth by Adam Smith.

The formal study of economics began when Adam Smith (1723–1790) published his famous book, *The Wealth of Nations*, in 1776. Many authors had written about economics in the centuries before Smith, but he was the first to address the subject in a comprehensive way.



In the first chapter of the book, Smith introduces the idea of the **division of labor**, which means that the way a good or service is produced is divided into a number of tasks that are performed by different workers, instead of all the tasks being performed by the same person. To illustrate the division of labor, Smith counted how many tasks were involved in making a pin:

drawing out a piece of wire, cutting it to the right length, straightening it, putting a head on one end and a point on the other, packaging pins for sale, and so on. Smith counted *eighteen* distinct tasks that were typically performed by different people—all for a pin!

Modern companies divide tasks, too. Even a relatively simple business like a restaurant divides up the task of serving meals into a range of jobs: top chef, sous chefs, less-skilled kitchen help, host/hostess, waiters/waitresses, janitors, a business manager to handle accounts and paychecks, etc. A complex business like a large manufacturing factory or a hospital can have hundreds of job classifications.

Why the Division of Labor Increases Production

When the tasks involved with producing a good or service are divided and subdivided, workers and businesses can produce a greater quantity of those goods or services. In his study of pin factories, Smith observed that one worker alone might make twenty pins in a day, but that a small business of ten workers (some of whom would need to do two or three of the eighteen tasks involved in pin making), could make forty-eight thousand pins in a day. How can a group of workers, each specializing in certain tasks, produce so much more than the same number of workers who try to produce the entire good or service by themselves? Smith offered three reasons.

First, **specialization** in a particular small job allows workers to focus on the parts of the production process in which they have an advantage. People have different skills, talents, and interests, so they will be better at some jobs than at others. The particular advantages may be based on educational choices, which are shaped, in turn, by interests and talents. Only those with medical training qualify to become doctors, for instance. For some goods, specialization will be affected by geography—it's easier to be a wheat farmer in North Dakota than in Florida, but easier to run a tourist hotel in Florida than in North Dakota. If you live in or near a big city, it's easier to attract enough customers to operate a successful dry-cleaning business or movie theater than if you live in a sparsely populated rural area. Whatever the reason, if people specialize in the production of what they do best, they will be more productive than if they produce a combination of things, some of which they are good at and some of which they are not.

Second, workers who specialize in certain tasks often learn to produce more quickly and with higher quality. This pattern holds true for many workers, including assembly-line laborers who build cars, stylists who cut hair, and doctors who perform heart surgery. In fact, specialized workers often know their jobs well enough to suggest innovative ways to do their work faster and better. A similar pattern often operates within businesses. In many cases, a business that focuses on one or a few products is more successful than firms that try to make a wide range of products.

Third, specialization allows businesses to take advantage of **economies of scale**, which means that, for many goods, as the level of production increases, the average cost of producing each individual unit declines. For example, if a factory produces only one hundred cars per year, each car will be quite expensive to make on average. However, if a factory produces fifty thousand cars each year, then it can set up an assembly line with huge machines and workers performing specialized tasks, and the average cost of production per car will drop. Economies of scale implies that production is becoming more efficient as the scale of production rises.

The ultimate result of workers who can focus on their preferences and talents, learn to do their specialized jobs better, and work in larger organizations is that society as a whole can produce and consume far more than if each person tried to produce all of their own goods and services. The division and specialization of labor has been a force against the problem of scarcity.

Trade and Markets

Specialization only makes sense, though, if workers (and other economic agents such as businesses and nations) can use their income to purchase the other goods and services they need. In short, specialization requires trade. You do not have to know anything about electronics or sound systems to play music—you just buy an iPod or MP3 player, download the music, and listen. You don't have to know anything about textiles or the construction of sewing machines if you need a jacket—you just buy the jacket and wear it. Instead of trying to acquire all the knowledge and skills involved in producing all of the goods and services that you wish to consume, the market allows you to learn a specialized set of skills and then use the pay you receive to buy the goods and services you need or want. This is how our modern society has evolved into a strong economy.

11. Reading: Microeconomics and Macroeconomics

Micro vs. Macro

It should be clear by now that economics covers a lot of ground. That ground can be divided into two parts: **Microeconomics** focuses on the actions of individual agents within the economy, like households, workers, and businesses; **macroeconomics** looks at the economy as a whole. It focuses on broad issues such as growth, unemployment, inflation, and trade balance. Microeconomics and macroeconomics are not separate subjects but are, rather, complementary perspectives on the overall subject of the economy.

To understand why both microeconomic and macroeconomic perspectives are useful, consider the problem of studying a biological ecosystem like a lake. One person who sets out to study the lake might focus on specific topics: certain kinds of algae or plant life; the characteristics of particular fish or snails; or the trees surrounding the lake. Another person might take an overall view and instead consider the entire ecosystem of the lake from top to bottom: what eats what, how the system remains in balance, and what environmental stresses affect this balance. Both approaches are useful, and both researchers study the same lake, but the viewpoints are different. In a similar way, both microeconomics and macroeconomics study the same economy, but each has a different starting point, perspective, and focus.

Whether you are looking at lakes or economics, the micro and the macro insights should illuminate each other. In studying a lake, the "micro" insights about particular plants and animals help us to understand the overall food chain, while the "macro" insights about the overall food chain help to explain the environment in which individual plants and animals live.

In economics, the micro decisions of individual businesses are influenced by the health of the macroeconomy—for example, firms will be more likely to hire workers if the overall economy is growing. In turn, the performance of the macroeconomy ultimately depends on the microeconomic decisions made by individual households and businesses.

Microeconomics

What determines how households and individuals spend their budgets? What combination of goods and services will best fit their needs and wants, given the budget they have to spend? How do people decide whether to work, and if so, whether to work full time or part time? How do people decide how much to save for the future, or whether they should borrow to spend beyond their current means?

What determines the products, and how many of each, a firm will produce and sell? What determines what prices a firm will charge? What determines how a firm will produce its products? What determines how many workers it will hire? How will a firm finance its business? When will a firm decide to expand, downsize, or even close? In the microeconomic part of this text, we will learn about the theory of consumer behavior and the theory of the firm.

Macroeconomics

What determines the level of economic activity in a society or nation?—that is, how many goods and services does it actually produce? What determines how many jobs are available in an economy? What determines a nation's standard of living? What causes the economy to speed up or slow down? What causes firms to hire more workers or lay them off? Finally, what causes the economy to grow over the long term?

An economy's macroeconomic health can be assessed by a number of standards or goals. The most important macroeconomic goals are the following:

- Growth in the standard of living
- Low unemployment
- Low inflation

Macroeconomic policy pursues these goals through monetary policy and fiscal policy:

- **Monetary policy**, which involves policies that affect bank lending, interest rates, and financial capital markets, is conducted by a nation's central bank. For the United States, this is the Federal Reserve.
- **Fiscal policy**, which involves government spending and taxes,

is determined by a nation's legislative body. For the United States, this is the Congress and the executive branch, which establishes the federal budget.

To keep the differences between these policies straight, remember that the term *monetary* relates to money, and the term *fiscal* relates to government revenue or taxes.

These are the main tools the government has to work with. Americans tend to expect that government can fix whatever economic problems we encounter, but to what extent is that expectation realistic? These are just some of the issues that will be explored later in this course.

Micro vs. Macro Video

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=30#oembed-1

Self Check: What Is Economics?

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the four Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1545

12. Outcome: Economic Models

What you'll learn to do: explain how economists use economic models

Earlier we defined economics as the study of how people choose to use scarce resources to best satisfy their unlimited wants. Economists try to analyze these choices both at the individual level (What wage does Alissa require in order to take that job?) and broader societal level (What is the impact of minimum wage on the unemployment rate?). The principal tool economists use to do this is models.

In this section, we'll define models and show how economists use them.

The specific things you'll learn in this section include the following:

- Define an economic model
- Explain why economic models are useful to economists
- Identify common economic models

Learning Activities

The learning activities for this section include the following:

- Reading: Using Economic Models

13. Reading: Using Economic Models

Economic Models and Math

Economists use models as the primary tool for explaining or making predictions about economic issues and problems. For example, an economist might try to explain what caused the Great Recession in 2008, or she might try to predict how a personal income tax cut would affect automobile purchases.

Economic models can be represented using words or using mathematics. All of the important concepts in this course can be explained without math. That said, math is a tool that can be used to explore economic concepts in very helpful ways. You know the saying "A picture is worth a thousand words"? The same applies to graphs: they're a very effective means of conveying information visually—without a thousand words. In addition to being a "picture," a graph is also a math-based model. Math is one way of working with (or manipulating) economic models.

Why would an economist use math when there are other ways of representing models, such as with text or narrative? Why would you use your fist to bang a nail, if you had a hammer? Math has certain advantages over text. It disciplines our thinking by making us specify exactly what we mean. You can get away with fuzzy thinking and vague approximations in your own mind, but not when you're reducing a model to algebraic equations. At the same time, math has certain disadvantages. Mathematical models lack the nuances that can be found in narrative models. The point is that math is one tool, but it's not the only tool or even always the best tool economists can use.

Examples of Models

An architect who is designing a major office building will probably build a physical model that sits on a tabletop to show how the entire city block will look after the new building is constructed. Companies often build models of their new products that are rougher and less finished than the final product but can still demonstrate how the new product will work and look. Such models help people visualize a product (or a building) in a more complete, concrete way than they could without them.

Similarly, economic models offer a way to get a complete view or picture of an economic situation and understand how economic factors fit together.

A good model to start with in economics is the **circular flow diagram** (Figure 1, below). Such a diagram indicates that the economy consists of two groups, households and firms, which interact in two markets: the **goods-and-services market** (also called the *product market*), in which firms sell and households buy, and the **labor market**, in which households sell labor to business firms or other employees.

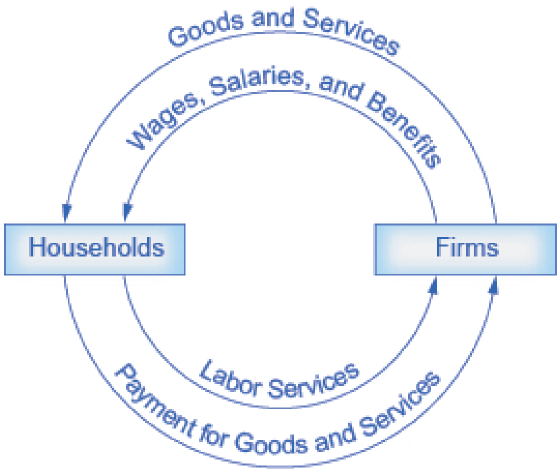


Figure 1. The Circular Flow Diagram

Of course, in the real world, there are many different markets for goods and services and markets for many different types of labor. The circular flow diagram simplifies these distinctions in order to make the picture easier to grasp. In the diagram, firms produce goods and services, which they sell to households in return for payments. The outer ring represents the two sides of the product market (which provides goods and services), in which households demand and firms supply. In addition, households (as workers) sell their labor to firms in return for wages, salaries, and benefits. This is shown in the inner circle, which represents the two sides of the labor market, in which households supply and firms demand. This version of the circular flow model is stripped down to the essentials, but it has enough features to explain how the product and labor markets work in the economy.

We could easily add details to this basic model if we wanted to introduce more real-world elements, like financial markets, governments, or interactions with the rest of the world (imports and exports). Economists reach for theories in much the same way as a carpenter might grab a tool. When economists identify an economic issue or problem, they sift through the available theories to see if they can find one that fits. Then they use the theory to give them insights about the issue or problem. In economics, theories are expressed as diagrams, graphs, or even as mathematical equations. Counter to what you might expect, economists don't figure out the solution to a problem and then draw the graph. Instead, they use the graph to help them discover the answer. At the introductory level, you can sometimes figure out the right answer without using a model, but if you keep studying economics, before too long you'll encounter issues and problems whose solution will require graphs. Both micro and macroeconomics are explained in terms of theories and models. The most well-known theories are probably those of supply and demand, but you will learn about several others.

Economic Models and Theories Video

An **economic model** is a simplified version of reality that allows us to observe, understand, and make predictions about economic behavior. The purpose of a model is to take a complex, real-world situation and pare it down to the essentials. If designed well, a model can give the analyst a better understanding of the situation and any related problems.

A good model is simple enough to be understood while complex enough to capture key information. Sometimes economists use the term *theory* instead of *model*. Strictly speaking, a theory is a more abstract representation, while a model is a more applied or empirical representation. Often, models are used to test theories. In this course, however, we will use the terms interchangeably.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=32#oembed-1

Self Check: Economic Models

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1546

14. Outcome: Math in Economics

What you'll learn to do: use mathematics in common economic applications

Economists use math as a tool for manipulating and exploring economic models. Sometimes it makes sense to express economic ideas in words; other times, math does a better job. What I would like to see is a statement that models can be articulated in different ways from verbal to mathematical. Economics is not math, but rather math is a tool for presenting and manipulating/exploring/using economic models. Many economic models use math to explain cause and effect. Don't worry, though: We're going to cover all the math you need to solve the problems in this course.

This section provides a quick review of some basic math (so you can avoid common errors) and then introduces the mathematical concepts you'll need throughout the course. Don't forget: return to this section later on if you encounter math that you can't follow.

Some students, when they find out that economics involves math, fear that the math will trip them up and prevent their success in the course—"I'm not a math person!" they think. If you share these thoughts, it may surprise you to know there's scientific research showing that when you practice new ways of thinking, your brain physically changes and adapts. Essentially, there's no such thing as "a math person" (or an "economics person"). You don't need a special talent or aptitude. It's mainly a matter of practice, hard work, and training your brain. The more you challenge your mind to learn, the more your brain cells connect to one another and the stronger those connections become.

So, how do you actually develop your brain and succeed in this kind of course? We asked some former students to tell us their advice for success. They said it was most important to

- Do the practice problems in the course.
- Ask questions.
- Study your mistakes.
- Explain the ideas to yourself in different ways until they are clear.

You will need to learn new things to pass this class, and you should expect it to feel hard as you wrestle with unfamiliar ideas and new ways of thinking. Don't give up, though! The feeling of struggling is a normal part of how the brain gets stronger when it learns things.

The specific things you'll learn in this section include the following:

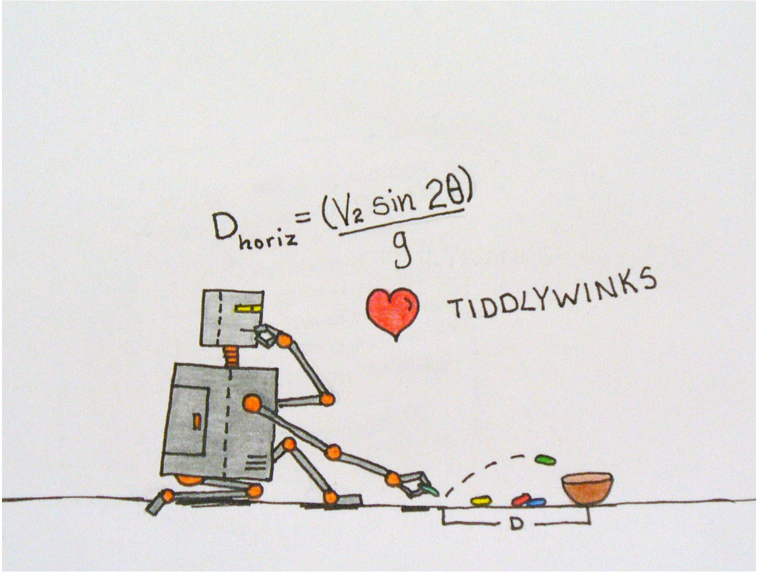
- Use an equation to understand the impact of changing economic variables
- Explain how functions describe cause and effect

Learning Activities

The learning activities for this section include the following:

- Reading: Purpose of Functions
- Reading: Solving Simple Equations

15. Reading: Purpose of Functions



Often, economic models are expressed in terms of mathematical functions. What's a function? Basically, a **function** describes a relationship involving one or more variables. Sometimes the relationship is a definition. For example (using words), Joan of Arc is a professor. This could be expressed as Joan of Arc = professor. Or, food = cherries, cheese, and chocolate means that cherries, cheese, and chocolate are food.

In economics, functions frequently describe cause and effect. The variable on the left-hand side is what is being explained ("the effect"). On the right-hand side is what's doing the explaining ("the causes"). Functions are also useful for making predictions. For example, think about your grade in this course. We might be able to predict how well you will do in this course by considering how well you've done in other courses, by how much you attend class or participate in the online activities, and by how many hours you study.

Not all of those things will have equal impact on your grade. Let's assume that your study time is most important and will have twice as much impact as the other factors. We are trying to describe 100 percent of the impact, so study time will explain 50 percent, attendance and participation will explain 25 percent, and your prior class grades will describe 25 percent. Together, this adds up to 100 percent.

Now, let's turn that into a function. Your grade in the course can be represented as the following:

Grade = (0.50 x hours_spent_studying) + (0.25 x class_attendance) + (0.25 x prior_GPA)

This equation states that your grade depends on three things: the number of hours you spend studying, your class attendance, and your prior course grades represented as your grade-point average (GPA). It also says that study time is twice as important (0.50) as either class_attendance (0.25) or prior_GPA score (0.25). If this relationship is true, how could you raise your grade in this course? By not skipping class and studying more. Note that you cannot do anything about your prior GPA, since that is calculated from courses you've already taken and grades you've already received.

Economic models tend to express relationships using economic variables, such as Budget = money_spent_on_econ_books + money_spent_on_music (assuming that the only things you buy are economics books and music). Often, there is some assumption that has to be explained in order to identify where the model has been simplified.

As you can see, in economic models the math isn't difficult. It's used to help describe and explain the relationships between variables.

16. Reading: Solving Simple Equations

Let's quickly review some math concepts that will help you avoid simple errors in your work.

Order of Operations

Remember, when you solve an equation it's important to do each operation in the following order:

1. Simplify inside parentheses and brackets.
2. Simplify the exponent.
3. Multiply and divide from left to right.
4. Add and subtract from left to right.

In this course you will not use exponents, but you will need to remember the order of the other steps.

So, in solving the following equation, you multiply first, then add:

$$y = 9 + 3 \times 10$$

$$y = 9 + 30$$

$$y = 39$$

Lines

In this course the most common equation you will see is $y = b + mx$. This is the equation for a line. We will revisit this equation later in this module when we review graphs. For now, let's practice solving this common equation using different variables.

Understanding Variables

To a mathematician or an economist, a **variable** is the name given to a quantity that can assume a range of values. In other words, the value of a variable can change or vary. In an equation it's represented by a letter or a symbol.

Because economic models often consider cause and effect, variables are important. You will often be asked to consider a range of options that result from different variables. Below is a very simple example:

$$y = 9 + 3x$$

In order to understand the range of options, we might start with 0. What does *y* equal if $x = 0$?

$$y = 9 + 3x$$
$$y = 9 + 3(0)$$
$$y = 9 + 0$$
$$y = 9$$

Now, let's look at the same formula with different information. What does y equal if $x = 5$?

$$y = 9 + 3x$$
$$y = 9 + 3(5)$$
$$y = 9 + 15$$
$$y = 24$$

Working with Variables

Remember that when you're trying to solve an equation with one or more variables, you need to isolate the variable. Let's walk through a simple example using the same equation from above.

What if we want to solve the equation in a case where $y = 24$?

$$y = 9 + 3x$$

$$24 = 9 + 3x$$

First, subtract the same number from each side of the equation to simplify the equation without changing the fact that it's an equality. In this case, we want to subtract the number that will enable us to isolate x (x is on one side of the equal sign all by itself). We can do that by subtracting 9 from each side.

$$24 = 9 + 3x$$

$$-9 = -9$$

$$15 = 3x$$

Now we can further simplify the equation by dividing both sides by 3.

$$\underline{15} = \underline{3x}$$

$$3 \qquad 3$$

$$5 = x$$

Let's practice solving for x one more time. What does x equal if $y = 12$?

$12 = 9 + 3x - 9 = -9 \quad 3 = 3x \quad \underline{3} = \underline{3x} \quad 3 \qquad 3 \quad 1 = x$

Self Check: Math in Economics

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1547

17. Outcome: Graphs in Economics

What you'll learn to do: use graphs in common economic applications

In this course, the most common way you will encounter economic models is in graphical form.

A graph is a visual representation of numerical information. Graphs condense detailed numerical information to make it easier to see patterns (such as "trends") among data. For example, which countries have larger or smaller populations? A careful reader could examine a long list of numbers representing the populations of many countries, but with more than two hundred nations in the world, searching through such a list would take concentration and time. Putting these same numbers on a graph, listing them from highest to lowest, would reveal population patterns much more readily.

Economists use graphs not only as a compact and readable presentation of data, but also for visually representing relationships and connections—in other words, they function as models. As such, they can be used to answer questions. For example: How do increasing interest rates affect home sales? Graphing the results can help illuminate the answers.

This section provides an overview of graphing—just to make sure you're up to speed on the basics. It's important to feel comfortable with the way graphs work before using them to understand new concepts.

The specific things you'll learn in this section include the following:

- Explain how a graph shows the relationship between two variables
- Differentiate between a positive relationship and a negative relationship
- Interpret economic information on a graph

Learning Activities

The learning activities for this section include the following:

- Video: Graph Review
- Reading: Creating and Interpreting Graphs
- Reading: Interpreting Slope
- Reading: Types of Graphs

18. Video: Graph Review

NOTE: Around the two-minute mark, the narrator inadvertently says "indirect," rather than "inverse." This is corrected later in the video.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=37#oembed-1

19. Reading: Creating and Interpreting Graphs

It's important to know the terminology of graphs in order to understand and manipulate them. Let's begin with a visual representation of the terms (shown in Figure 1), and then we can discuss each one in greater detail.

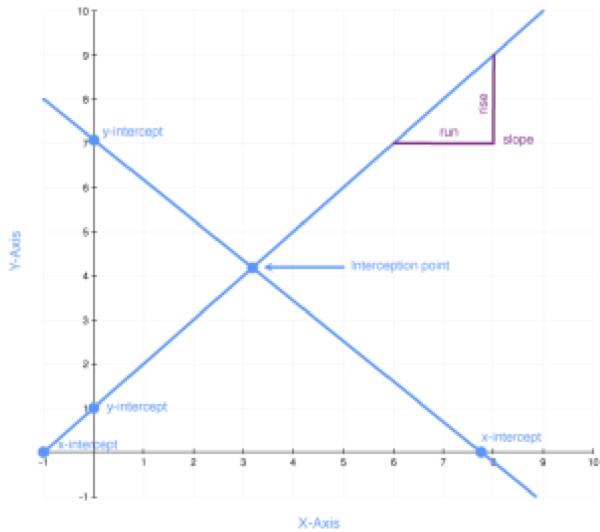


Figure 1. Graph Terminology

Throughout this course we will refer to the horizontal line on the graph as the **x-axis**. We will refer to the vertical line on the graph as the **y-axis**. This is the standard convention for graphs.

An **intercept** is where a line on a graph crosses ("intercepts") the x-axis or the y-axis. You can see the x-intercepts and y-intercepts on the graph above. The point where two lines on a graph cross is called the **interception point**.

The other important term to know is *slope*. The slope tells us how steep a line on a graph is. Technically, **slope** is the change in the vertical axis divided by the change in the horizontal axis. The formula for calculating the slope is often referred to as the "rise over the run"—again, the change in the distance on the y-axis (rise) divided by the change in the x-axis (run).

Now that you know the “parts” of a graph, let’s turn to the equation for a line:

“>y = b + mx

Let’s use the same equation we used earlier, in the section on solving algebraic equations:

y = 9 + 3x

“>In this equation for a line, the *b* term is 9 and the *m* term is 3. The ta

Values for the Slope Intercept Equation

x	**y**
0	9
1	12
2	15
3	18
4	21
5	24
6	27

Next we can place each of these points on a graph. We can start with 0 on the *x*-axis and plot a point at 9 on the *y*-axis. We can do the same with the other pairs of values and draw a line through all the points, as on the graph in Figure 2, below.

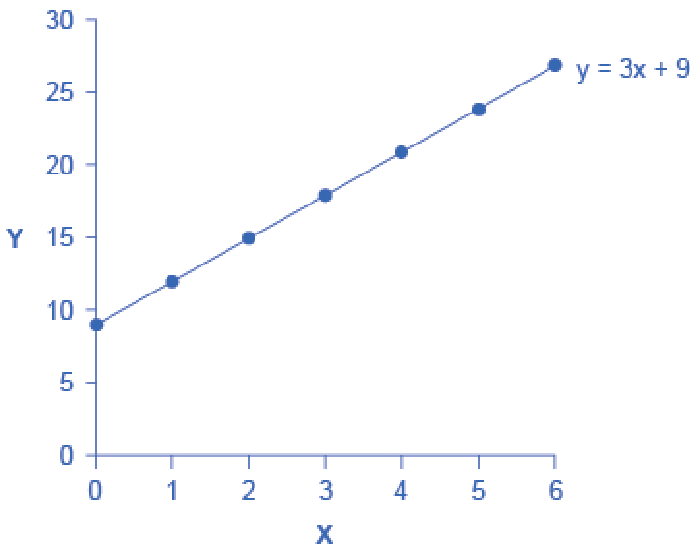


Figure 2. Slope and Algebra of a Straight Line

This example illustrates how the *b* and *m* terms in an equation for a straight line determine the shape of the line. The *b* term is called the *y*-intercept. The reason is that if $x = 0$, the *b* term will reveal where the line intercepts, or crosses, the y-axis. In this example, the line hits the vertical axis at 9. The m term in the equation for the line is the slope. Remember that slope is defined as rise over run; the slope of a line from one point to another is the change in the vertical axis divided by the change in the horizontal axis. In this example, each time the x term increases by 1 (the run), the y term rises by 3. Thus, the slope of this line is 3. Specifying a y-intercept and a slope—that is, specifying b and m in the equation for a line—will identify a specific line. Although it is rare for real-world data points to arrange themselves as a perfectly straight line, it often turns out that a straight line can offer a reasonable approximation of actual data.

20. Reading: Interpreting Slope

What the Slope Means

The concept of slope is very useful in economics, because it measures the relationship between two variables. A **positive slope** means that two variables are positively related—that is, when x increases, so does y, and when x decreases, y decreases also. Graphically, a positive slope means that as a line on the line graph moves from left to right, the line rises. We will learn in other sections that "price" and "quantity supplied" have a positive relationship; that is, firms will supply more when the price is higher.

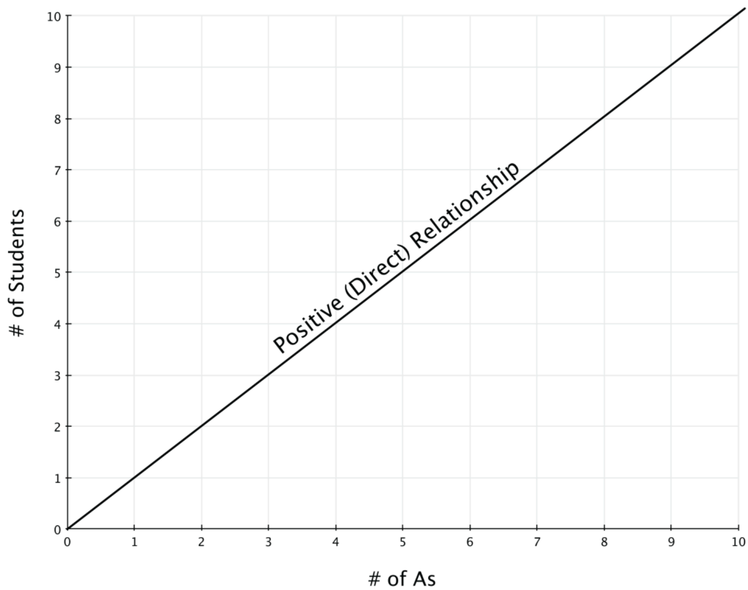


Figure 1. Positive Slope

A **negative slope** means that two variables are negatively related; that is, when x increases, y decreases, and when x decreases, y increases. Graphically, a negative slope means that as the line on the line graph moves from left to right, the line falls. We will learn that "price" and "quantity demanded" have a negative relationship; that is, consumers will purchase less when the price is higher.

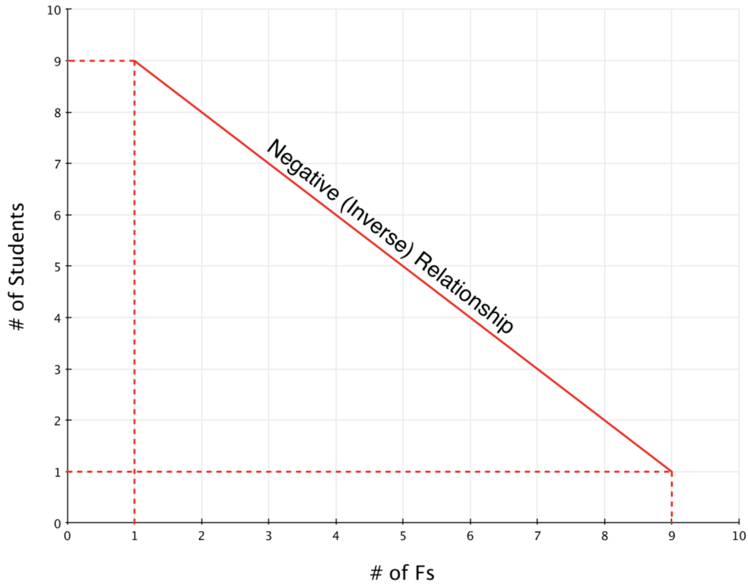


Figure 2. Negative slope

A **slope of zero** means that there is a constant relationship between *x* and *y*. Graphically, the line is flat; the rise over run is zero.

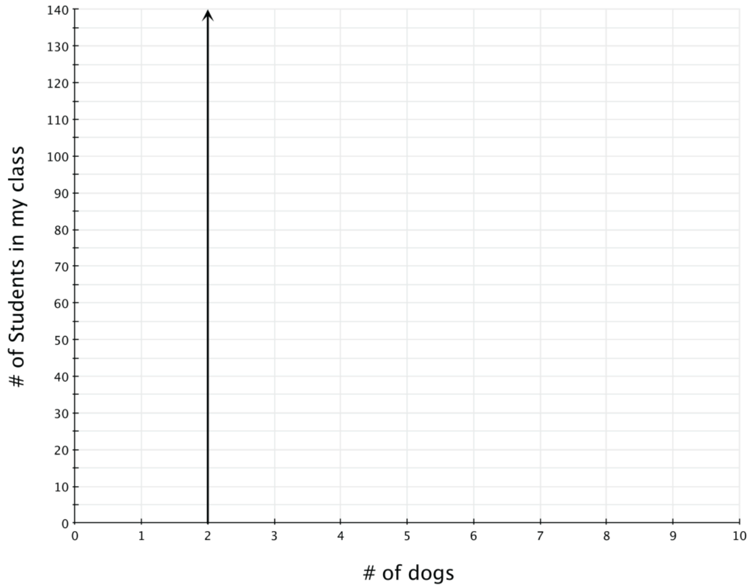


Figure 3. Slope of Zero

The unemployment-rate graph in Figure 4, below, illustrates a common pattern of many line graphs: some segments where the slope is positive, other segments where the slope is negative, and still other segments where the slope is close to zero.

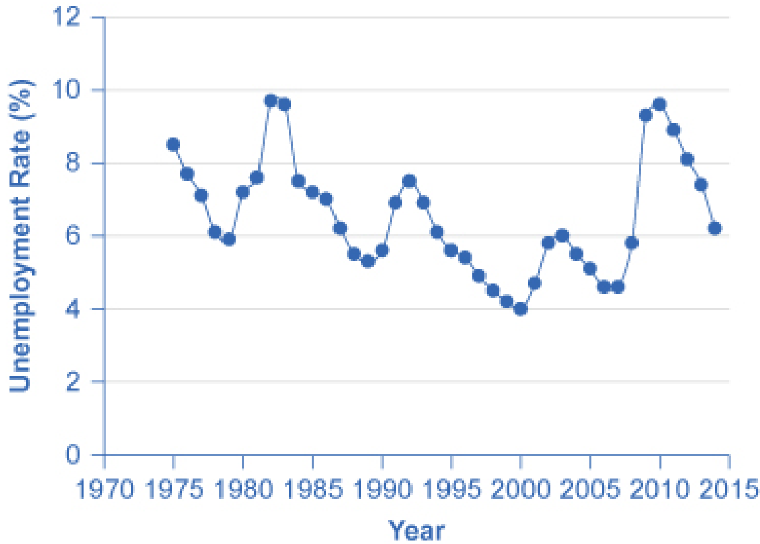


Figure 4. U.S. Unemployment Rate, 1975–2014

Calculating Slope

The slope of a straight line between two points can be calculated in numerical terms. To calculate slope, begin by designating one point as the "starting point" and the other point as the "end point" and then calculating the rise over run between these two points.

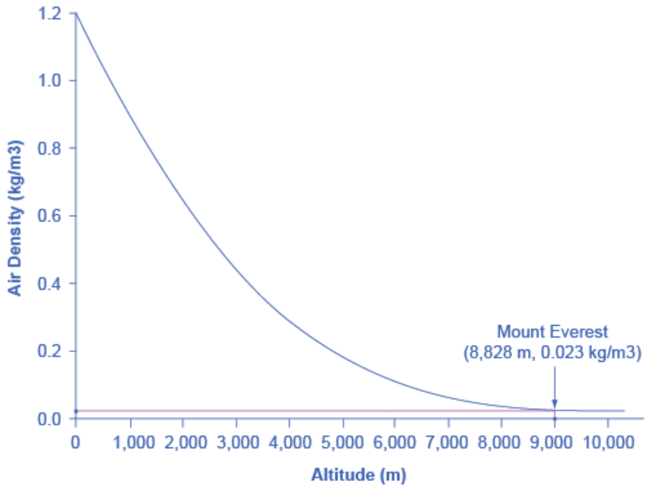


Figure 5. Altitude–Air Density Relationship

As an example, consider the slope of the air-density graph, above, between the points representing an altitude of 4,000 meters and an altitude of 6,000 meters:

Rise: Change in variable on vertical axis (end point minus original point)

“>0.100 – 0.307 = –0.207

Run: Change in variable on horizontal axis (end point minus original point)

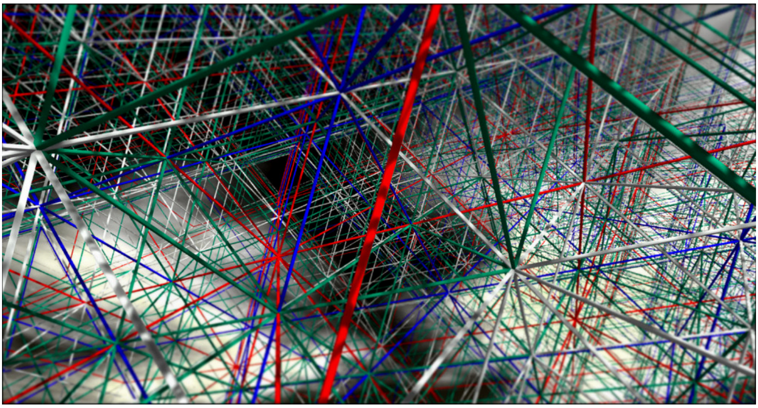
“>6,000 – 4,000= 2,000

Thus, the slope of a straight line between these two points would be the following: from the altitude of 4,000 meters up to 6,000 meters, the density of the air decreases by approximately 0.1 kilograms/cubic meter for each of the next 1,000 meters.

Suppose the slope of a line were to increase. Graphically, that means it would get steeper. Suppose the slope of a line were to decrease. Then it would get flatter. These conditions are true whether or not the slope was positive or negative to begin with. A higher positive slope means a steeper upward tilt to the line, while a smaller positive slope means a flatter upward tilt to the line. A negative slope that is larger in absolute value (that is, more negative) means a steeper downward tilt to the line. A slope of zero is a horizontal flat line. A vertical line has an infinite slope.

Suppose a line has a larger intercept. Graphically, that means it would shift out (or up) from the old origin, parallel to the old line. If a line has a smaller intercept, it would shift in (or down), parallel to the old line.

21. Reading: Types of Graphs



Three types of graphs are used in this course: line graphs, pie graphs, and bar graphs. Each is discussed below.

Line Graphs

The graphs we've discussed so far are called line graphs, because they show a relationship between two variables: one measured on the horizontal axis and the other measured on the vertical axis.

Sometimes it's useful to show more than one set of data on the same axes. The data in the table, below, is displayed in Figure 1, which shows the relationship between two variables: length and median weight for American baby boys and girls during the first three years of life. (The median means that half of all babies weigh more than this and half weigh less.) The line graph measures length in inches on the horizontal axis and weight in pounds on the vertical axis. For example, point A on the figure shows that a boy who is

28 inches long will have a median weight of about 19 pounds. One line on the graph shows the length-weight relationship for boys, and the other line shows the relationship for girls. This kind of graph is widely used by health-care providers to check whether a child's physical development is roughly on track.

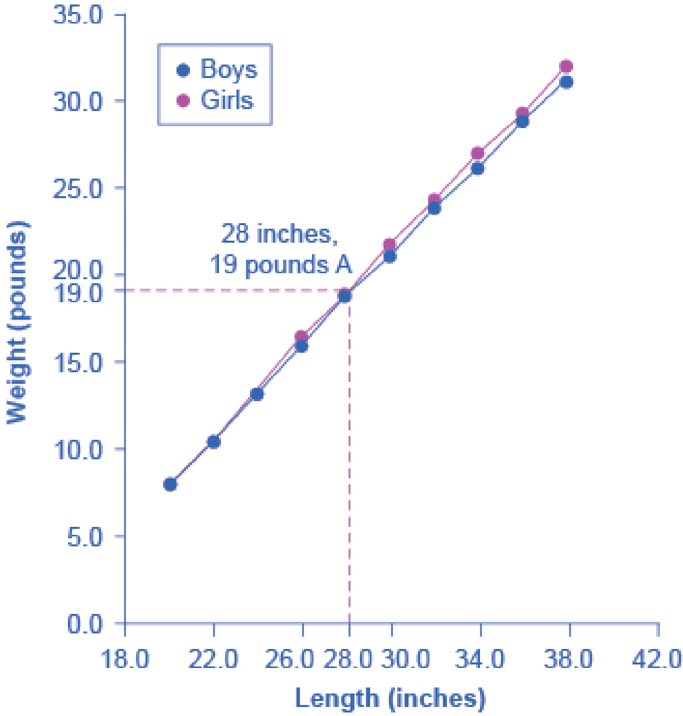


Figure 1. The Length-Weight Relationship for American Boys and Girls

Length-to-Weight Relationship for American Boys and Girls

Boys from Birth to 36 Months		**Girls from Birth to 36 Months**	
Length (inches)	Weight (pounds)	Length (inches)	Weight (pounds)
20.0	8.0	20.0	7.9
22.0	10.5	22.0	10.5
24.0	13.5	24.0	13.2
26.0	16.4	26.0	16.0
28.0	19.0	28.0	18.8
30.0	21.8	30.0	21.2
32.0	24.3	32.0	24.0
34.0	27.0	34.0	26.2
36.0	29.3	36.0	28.9
38.0	32.0	38.0	31.3

Not all relationships in economics are linear. Sometimes they are curves. Figure 2, below, presents another example of a line graph, representing the data from the table underneath. In this case, the line graph shows how thin the air becomes when you climb a mountain. The horizontal axis of the figure shows altitude, measured in meters above sea level. The vertical axis measures the density of the air at each altitude. Air density is measured by the weight of the air in a cubic meter of space (that is, a box measuring one meter in height, width, and depth). As the graph shows, air pressure is heaviest at ground level and becomes lighter as you climb. Figure 1 shows that a cubic meter of air at an altitude of 500 meters weighs approximately one kilogram (about 2.2 pounds). However, as the altitude increases, air density decreases. A cubic meter of air at the top of Mount Everest, at about 8,828 meters, would weigh only 0.023 kilograms. The thin air at high altitudes explains why many mountain climbers need to use oxygen tanks as they reach the top of a mountain.

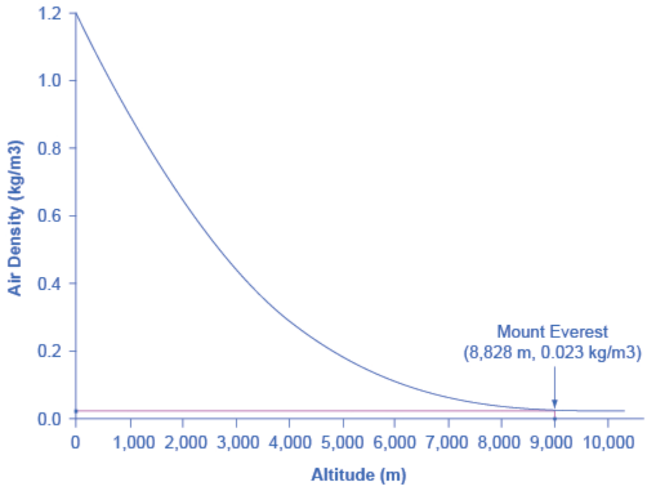


Figure 2. Altitude–Air-Density Relationship

Altitude–to–Air-Density Relationship

Altitude (meters)	Air Density (kg/cubic meters)
0	1.200
500	1.093
1,000	0.831
1,500	0.678
2,000	0.569
2,500	0.484
3,000	0.415
3,500	0.357
4,000	0.307
4,500	0.231
5,000	0.182
5,500	0.142
6,000	0.100
6,500	0.085
7,000	0.066
7,500	0.051
8,000	0.041
8,500	0.025
9,000	0.022
9,500	0.019
10,000	0.014

The length-weight relationship and the altitude–air-density relationship in these two figures represent averages. If you were to collect actual data on air pressure at different altitudes, the same altitude in different geographic locations would have slightly different air density, depending on factors like how far you were from the equator, local weather conditions, and the humidity in the air. Similarly, in measuring the height and weight of children for the previous line graph, children of a particular height would have a range of different weights, some above average and some below. In the real world, this sort of variation in data is common. The task of a researcher is to organize that data in a way that helps to understand typical patterns. The study of statistics, especially when combined with computer statistics and spreadsheet programs, is a great help in organizing this kind of data, plotting line graphs, and looking for typical underlying relationships. For most economics and social science majors, a statistics course will be required at some point.

One common line graph is called a time series, in which the horizontal axis shows time and the vertical axis displays another variable. Thus, a time-series graph shows how a variable changes over time. Figure 3 shows the unemployment rate in the United States since 1975, where unemployment is defined as the percentage of adults who want jobs and are looking for a job, but cannot find one. The points for the unemployment rate in each year are plotted on the graph, and a line then connects the points, showing how the unemployment rate has moved up and down since 1975. With a graph like this, it is easy to spot the times of high unemployment and of low unemployment.

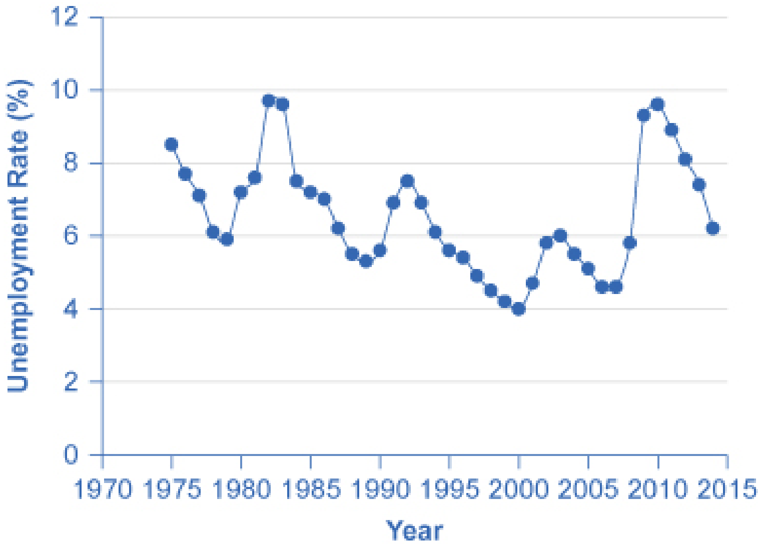


Figure 3. U.S. Unemployment Rate, 1975–2014

Pie Graphs

A pie graph (sometimes called a pie chart) is used to show how an overall total is divided into parts. A circle represents a group as a whole. The slices of this circular "pie" show the relative sizes of subgroups.

Figure 4 shows how the U.S. population was divided among children, working-age adults, and the elderly in 1970, 2000, and what is projected for 2030. The information is first conveyed with numbers in the table, below, and then in three pie charts.

U.S. Age Distribution, 1970, 2000, and 2030 (projected)

Year	Total Population	19 and Under	20–64 years	Over 65
1970	205.0 million	77.2 (37.6%)	107.7 (52.5%)	20.1 (9.8%)
2000	275.4 million	78.4 (28.5%)	162.2 (58.9%)	34.8 (12.6%)
2030	351.1 million	92.6 (26.4%)	188.2 (53.6%)	70.3 (20.0%)

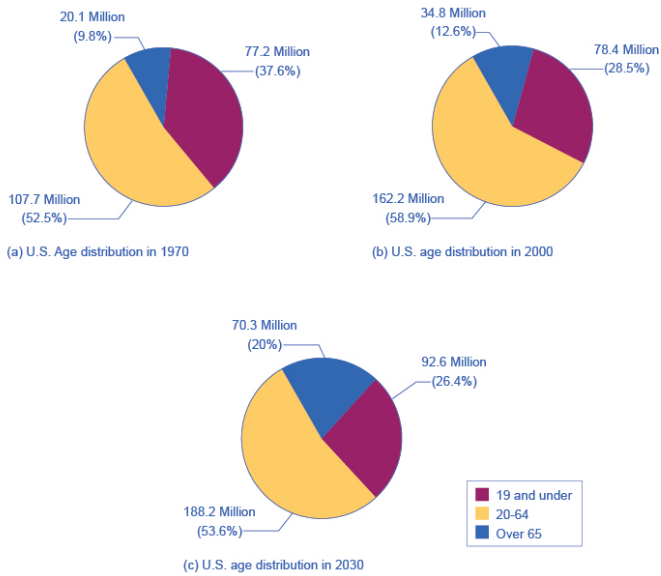


Figure 4. Pie Graphs of the U.S. Age Distribution (numbers in millions)

In a pie graph, each slice of the pie represents a share of the total, or a percentage. For example, 50% would be half of the pie and 20% would be one-fifth of the pie. The three pie graphs in Figure 4 show that the share of the U.S. population 65 and over is growing. The pie graphs allow you to get a feel for the relative size of the different age groups from 1970 to 2000 to 2030, without requiring you to slog through the specific numbers and percentages in the table. Some common examples of how pie graphs are used include dividing the population into groups by age, income level, ethnicity, religion, occupation; dividing different firms into categories by size, industry, number of employees; and dividing up government spending or taxes into its main categories.

Bar Graphs

A bar graph uses the height of different bars to compare quantities. The table, below, lists the 12 most populous countries in the world. Figure 5 provides this same data in a bar graph. The height of the bars corresponds to the population of each country. Although you may know that China and India are the most populous countries in the world, seeing how the bars on the graph tower over the other countries helps illustrate the magnitude of the difference between the sizes of national populations.

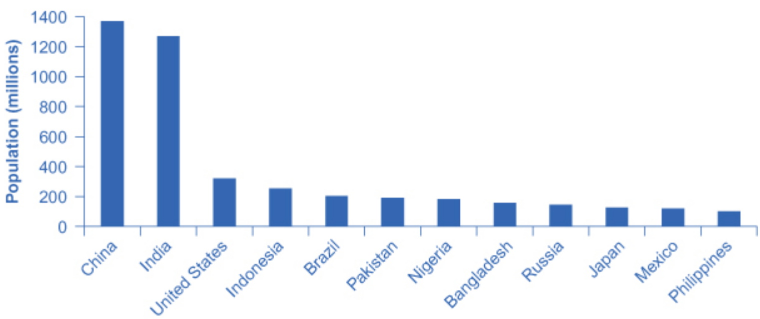


Figure 5. Leading Countries of the World by Population, 2015 (in millions)

Leading 12 Countries of the World by Population

Country	Population
China	1,369
India	1,270
United States	321
Indonesia	255
Brazil	204
Pakistan	190
Nigeria	184
Bangladesh	158
Russia	146
Japan	127
Mexico	121
Philippines	101

Bar graphs can be subdivided in a way that reveals information similar to that we can get from pie charts. Figure 6 offers three bar graphs based on the information from Figure 4 about the U.S. age distribution in 1970, 2000, and 2030. Figure 6 (a) shows three bars for each year, representing the total number of persons in each age bracket for each year. Figure 6 (b) shows just one bar for each year, but the different age groups are now shaded inside the bar. In Figure 6 (c), still based on the same data, the vertical axis measures percentages rather than the number of persons. In this case, all three bar graphs are the same height, representing 100 percent of the population, with each bar divided according to the percentage of population in each age group. It is sometimes easier for a reader to run his or her eyes across several bar graphs, comparing the shaded areas, rather than trying to compare several pie graphs.

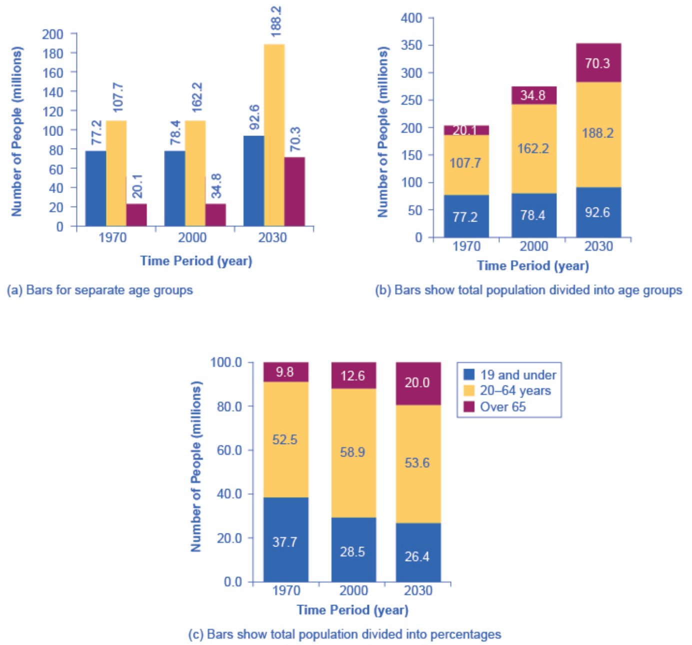


Figure 6. U.S. Population with Bar Graphs

Figure 5 and Figure 6 show how the bars can represent countries or years, and how the vertical axis can represent a numerical or a percentage value. Bar graphs can also compare size, quantity, rates, distances, and other quantitative categories.

Comparing Line Graphs, Pie Charts, and Bar Graphs

Now that you are familiar with pie graphs, bar graphs, and line graphs, how do you know which graph to use for your data? Pie graphs are often better than line graphs at showing how an overall group is divided. However, if a pie graph has too many slices, it can become difficult to interpret.

Bar graphs are especially useful when comparing quantities. For example, if you are studying the populations of different countries, as in Figure 5, bar graphs can show the relationships between the population sizes of multiple countries. Not only can it show these relationships, but it can also show breakdowns of different groups within the population.

A line graph is often the most effective format for illustrating a relationship between two variables that are both changing. For example, time-series graphs can show patterns as time changes, like the unemployment rate over time. Line graphs are widely used in economics to present continuous data about prices, wages, quantities bought and sold, the size of the economy.

Self Check: Graphs in Economics

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the three Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1548

22. Problem Set: Math in Economics

Test your understanding of the learning outcomes in this module by working through the following problems. These problems aren't graded, but they give you a chance to practice before taking the quiz.

If you'd like to try a problem again, you can click the link that reads, "Try another version of this question."

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=41

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=41

One or more interactive elements has been excluded from this version of the text. You can view them online *here:* *https://library.achievingthedream.org/sacmicroeconomics/?p=41*

One or more interactive elements has been excluded from this version of the text. You can view them online here: *https://library.achievingthedream.org/sacmicroeconomics/?p=41*

23. Problem Set: Graphs in Economics

Test your understanding of the learning outcomes in this module by working through the following problems. These problems aren't graded, but they give you a chance to practice before taking the quiz.

If you'd like to try a problem again, you can click the link that reads, "Try another version of these questions."

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=42

24. Putting It Together: Economic Thinking

Summary

The goal of this module was to introduce you to the basic questions and tools of economics. You learned how to:

- Explain what economics is and why it is important
- Explain how economists use economic models
- Use mathematics in common economic applications
- Use graphs in common economic applications

The Cost of Waiting in Line

Given your new understanding of economic concepts, let's revisit the example at the beginning of this module: the experience of taking a flight and feeling like others have special privileges that you don't have.

In our example, you waited in line at the security checkpoint for much longer than those who went through the express line. Let's assume that you waited in line for one hour and 10 minutes, while those with express access were able to get through security in just 10 minutes. What is the cost of one hour of your time? What is the cost of a ticket that gets you into the the express lane? Did you make the right choice?

Let's return to the concept of opportunity cost. Remember, opportunity cost indicates what must be given up to obtain something that's desired.

You chose to wait in line rather than buying a ticket that would allow you to use the express lane. It may not have seemed like a choice, but you did choose to buy a less expensive ticket—instead of paying more for one would have gotten you into the express line. How much more money would you have to pay for that ticket?

Depart	Arrive	Flight #	Routing	Travel Time	Business Select $596 - $605	Anytime $568 - $577	Wanna Get Away $139 - $276
6:00 AM	3:45 PM	1415 1472	1 stop Change Planes DEN	6h 45m	$605	$577	$276

On a flight from Los Angeles to Baltimore, the Business Select fare is $605. This is the fare that permits access to the express security lane. The lowest fare is $276. The difference is:

$605 – $276 = $329

In other words, you chose to wait in line for one hour in order to save the $329 that you would have had to spend for a Business Select ticket. When we think of this in terms of opportunity cost, you now have some way to measure your decision: you have a firm number that can be compared against the cost of an alternative.

The Full Cost of Your Time

Let's consider monetary costs alone. If you had worked at a job for one more hour—instead of waiting in the security line—you could have earned an additional $20. That choice would have meant earning more money, but you would have had to spend far more for a Business Select ticket than you made in an hour. Since the $329 cost of that ticket is so much greater than the $20 you might have earned, the decision to wait in line for an hour (as opposed to working one hour more) makes good sense when comparing the monetary cost.

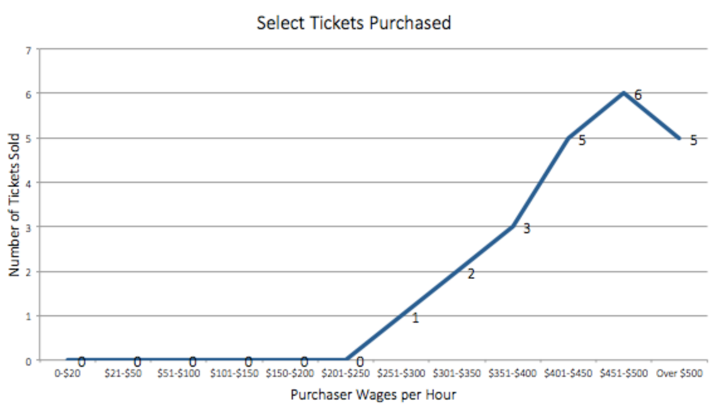
But remember, *opportunity cost is the value of the next best* *alternative*, and there are likely ways that you spend your time that you value more than money. Perhaps getting through security more quickly will enable you to sit down in a quiet café, enjoy a cup of tea, and avoid the stress you feel when you are rushed. There is a real value that you have given up.

Or, what if you are struggling to stay caught up on your schoolwork and an hour of study makes the difference between getting a good grade on an important test or not? If it's a test that has a big impact on your grade or academic record, then an hour might be incredibly valuable.

In other words, for any individual, the opportunity cost may simply be the lost money from work, or it may be peace of mind, or it may be an hour of study time—or something altogether different. Whichever one of the alternatives has the greatest value to you is your opportunity cost for one hour in line.

Does It Ever Make Sense to Buy Time?

Obviously an hour of time has a different opportunity cost for every individual. Let's take a minute to look at a graph of hourly wages and see if it helps us think about the opportunity cost.



On any flight, passengers with a higher hourly wage are more likely to purchase a Select ticket that permits them to bypass security lines and board early. For some groups of passengers, the time is more valuable than the money. Still, you will note that some passengers are willing to spend more than their hourly wage to avoid waiting in line.

The definition of opportunity cost is quite specific: it's the value of the next best alternative. However, every individual values time, money, convenience, peace of mind, and other factors differently.

25. Glossary: Economic Thinking

circular flow diagram a diagram indicating that the economy consists of households and firms interacting in a goods-and-services market and a labor market

direct relationship a relationship between two variables such that both either increase or decrease together; *also called a "positive relationship"*

division of labor the way in which the work required to produce a good or service is divided into tasks performed by different workers

economic model is a simplified version of reality that allows us to observe, understand, and make predictions about economic behavior

economics is the study of how humans make choices under conditions of scarcity in an attempt to satisfy their unlimited wants

economies of scale when the average cost of producing each individual unit declines as total output increases

function a relationship or expression involving one or more variables

goods-and-services market a market in which firms are sellers of what they produce, and households are buyers

independent relationship a relationship between two independent variables such that when one changes, the other does not change, and vice versa; *also called a "constant relationship"*

intercept the point on a graph where a line crosses the vertical axis or horizontal axis

interception point the point on a graph where two lines cross

inverse relationship a relationship between two variables such that when one increases, the other decreases, or vice versa; *also called "negative relationship"*

labor market the market in which households sell their labor as workers to businesses or other employers

macroeconomics the branch of economics that focuses on broad issues such as growth, unemployment, inflation, and trade balance

microeconomics the branch of economics that focuses on actions of particular agents within the economy, like households, workers, and businesses

monetary policy policy that involves altering the level of interest rates, the availability of credit in the economy, and the extent of borrowing

negative slope indicates that two variables are negatively related; when one variable increases, the other decreases, and when one variable decreases, the other increases

opportunity cost is the value of the next best alternative

positive slope indicates that two variables are positively related; when one variable increases, so does the other, and when one variable decreases, the other also decreases

productive resources the inputs used in the production of goods and services to make a profit: land, economic capital, labor, and entrepreneurship; *also called "factors of production"*

scarcity exists when human wants for goods and services exceed the available supply

slope the change in the vertical axis divided by the change in the horizontal axis

slope of zero indicates that there is a constant relationship between two variables: when one variable changes, the other does not change

specialization when workers or firms focus on particular tasks for which they are well suited within the overall production process

variable a quantity that can assume a range of values

x-axis the horizontal line on a graph

y-axis the vertical line on a graph

PART III

MODULE: CHOICE IN A WORLD OF SCARCITY

26. Why It Matters: Choice in a World of Scarcity

Why use economic thinking to explain choice in a world of scarcity?

As you now know, the study of economics is about choices that are made by individuals and entities, given the fact that we can never have enough. You might not argue that you don't have enough time or money, for instance, but why might you want to think about that in economic terms? Let's look at one situation in which the choices you make today—with limited time and money—have an impact on the choices available to you in the future.

It's generally true that the higher educational degree a person has, the higher the salary he or she will earn. So why aren't more people pursuing higher degrees? The short answer: choices and trade-offs.

In 2012, the annual salary for a full-time U.S. worker over age twenty-five with a master's degree was $67,600. Compare that to annual earnings for a full-time worker over twenty-five with no higher than a bachelor's degree: $55,432 a year. What about those with no higher than a high school diploma? They earn just $33,904 in a year. In other words, says the Bureau of Labor Statistics (BLS), earning a bachelor's degree boosted salaries 63 percent above what you would have earned if you had stopped your education after high school. A master's degree yields a salary almost double that of a high school diploma.

What are your educational goals? Do you plan to complete a bachelor's degree? A master's degree? Given the salary data, shouldn't everyone pursue a master's degree? When you made your own educational plans and goals, perhaps you were motivated by the potential for financial returns later on—i.e., the expectation that a higher degree would lead to a higher-paid job or career. But what other factors did you consider? Perhaps you also thought about the time and cost of education and the other things you like to do when you aren't studying. Other people, it turns out, also think about these things when deciding whether or not to pursue college.

Considering salary data alone, you might expect a lot of people to choose to attend college and at least earn a bachelor's degree. In fact, in 2012, the BLS reported that while nearly 88 percent of the U.S. population had a high school diploma, only 31 percent had a bachelor's degree, and only 8 percent had earned a master's degree.

For the majority of Americans, the time, money, and effort required to earn a degree is too great, in spite of the resulting salary benefits. In recognition of these barriers, state and federal governments have created programs such as the Pell Grant program to help students pay the financial costs of going to college. However, these programs don't cover the opportunity costs that are often the most pressing concern for students. For example, the opportunity cost of lost income that could be used to support a student's family might be a significant factor.

So, now that you're in college, how can you make the best decision about which level of education to pursue? Perhaps more important, how can you be realistic about your scarce resources and develop a plan that provides the greatest benefit to you?

In this module we will look more closely at the idea of choices and trade-offs, revisit the concept of opportunity cost, and learn how to calculate it. This will help you assign dollar amounts to your choices and understand why your decision to pursue a college degree—in spite of the opportunity costs—is one of the most important decisions you can make toward improving your financial future.

Learning Outcomes

- Explain the cost of choices and trade-offs
- Illustrate society's trade-offs by using a production possibilities frontier (or curve)
- Explain the assumption of economic rationality by individuals and firms
- Define marginal analysis
- Differentiate between positive and normative statements

27. Outcome: The Cost of Choices

What you'll learn to do: explain the cost of choices and trade-offs

In the previous module we introduced the concepts of scarcity, monetary cost, and opportunity cost. This section focuses on the actual calculation of those costs, which becomes important when you are trying to understand trade-offs in very concrete terms.

The specific things you'll learn in this section include the following:

- Explain how budget constraints impact choices
- Calculate the opportunity cost of an action

Learning Activities

The learning activities for this section include the following:

- Reading: Budget Constraints and Choices
- Reading: Calculating Opportunity Cost

28. Reading: Budget Constraints and Choices

Budget Constraint Framework

For most of us, the idea of scarcity and trade-offs is something we experience in a very real way when it comes to our own budget constraints. Most of us have a limited amount of money to spend on the things we need and want. Another kind of budget constraint is time. For instance, as a student, you only have twenty-four hours in the day to study, eat, sleep, and check Facebook. An hour spent studying economics is an hour that can't be used for sleep or play (or something else). As a result, you have to make choices and trade-offs.

In economics, a **budget constraint** refers to all possible combinations of goods that someone can afford, given the prices of goods, when all income (or time) is spent.

Take the following example of someone who must choose between two different goods: Charlie has $10 in spending money each week that he can allocate between bus tickets for getting to work and the burgers he eats for lunch. Burgers cost $2 each, and bus tickets are 50 cents each. Figure 1, below, shows Charlie's budget constraint ($10) and all the possible combinations of burgers and bus tickets he can afford if he spends all his money.

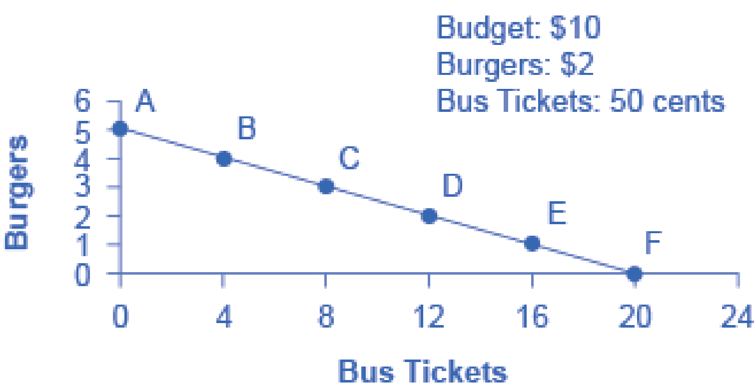


Figure 1. *Charlie's budget constraint*

The vertical axis in the figure shows burger purchases, and the horizontal axis shows bus ticket purchases. If Charlie spends all his money on burgers, he can afford five per week. ($10 per week/$2 per burger = 5 burgers per week.) But if he does this, he won't be able to afford any bus tickets. This choice (zero bus tickets and 5 burgers) is shown by point A in the figure. Alternatively, if Charlie spends all his money on bus tickets, he can afford 20 per week. ($10 per week/$0.50 per bus ticket = 20 bus tickets per week.) Then, however, he will not be able to afford any burgers. This alternative choice (20 bus tickets and zero burgers) is shown by point F. The slope of the budget constraint is determined by the relative price of burgers and bus tickets.

If Charlie is like most people, he will choose some combination that includes both bus tickets and burgers—that is, he will choose one of the points along the budget-constraint line that connects points A and F. Each point inside or on the budget constraint shows a combination of burgers and bus tickets that Charlie can afford. (A point inside the curve is definitely an option—it just means that Charlie isn't spending all his money.) Keep in mind that the curve represents the *maximum* number of burgers and bus tickets he can buy. Any point outside the constraint is not affordable, because it would cost more money than Charlie has in his budget.

The budget constraint clearly shows the trade-off Charlie faces in choosing between burgers and bus tickets. Suppose he is currently at point D, where he can afford 12 bus tickets and 2 burgers. What would it cost Charlie for one more burger? It would be natural to answer $2, but that's not the way economists think. Instead, they ask: How many bus tickets would Charlie have to give up to get one more burger, while staying within his budget? The answer is four bus tickets. That is the true cost to Charlie of one more burger.

A budget-constraint diagram like the one above, with just two goods—burgers and bus tickets—is simple and not very realistic. After all, in an economy like ours (and Charlie's), people choose from thousands of goods. However, economists use graphs and models to illustrate that every choice has an **opportunity cost**, which is the point that carries over to the real world.

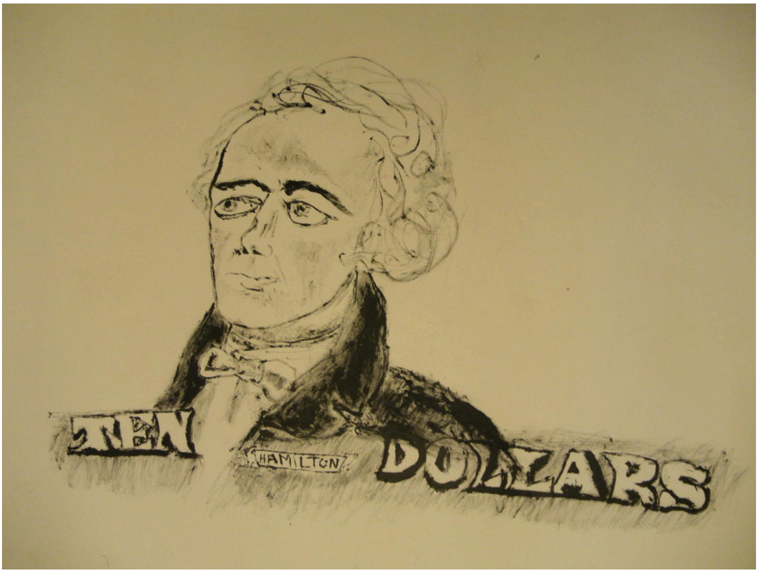
Sunk Costs

In the budget constraint framework, all decisions involve what will happen next: What quantities of goods will you consume? How many hours will you work? How much will you save? Choices made or costs in the past are not taken into account. The budget constraint framework assumes that **sunk costs**—costs incurred in the past that can't be recovered—should not affect the current decision.

Suppose you pay $8 to see a movie, but after watching the first thirty minutes, you decide that it's awful. Should you stick it out and watch the rest because you paid for the ticket, or should you leave? The money you spent on the ticket is a sunk cost, and unless the theater manager is feeling generous, you won't get a refund. But staying for the rest of the movie means paying an opportunity cost in time. Your choice is whether to spend the next ninety minutes suffering through a rotten movie or do something—anything—else. The lesson of sunk costs is to forget about the money and time that is irretrievably gone and to focus, instead, on the costs and benefits of current and future options. A sunk cost is water under the bridge, so to speak.

For people and organizations alike, dealing with sunk costs can be frustrating and difficult. For one thing, it often means admitting an earlier error of judgment. Many companies find it hard to give up on a new product that's doing poorly because they've invested so much time and money in the product development and launch. But the lesson of sunk costs is to ignore them and make decisions based on what will happen in the future.

29. Reading: Calculating Opportunity Cost



It makes intuitive sense that Charlie can buy only a limited number of bus tickets and burgers with a limited budget. Also, the more burgers he buys, the fewer bus tickets he can buy. With a simple example like this, it isn't too hard to determine what he can do with his very small budget, but when budgets and constraints are more complex, it's important to know how to solve equations that demonstrate budget constraints and opportunity cost.

Very simply, when Charlie is spending his full budget on burgers and tickets, his budget is equal to the total amount that he spends on burgers plus the total amount that he spends on bus tickets. For example, if Charlie buys four bus tickets and four burgers with his $10 budget (point B on the graph below), the equation would be

$$\$10 = (\$2 \times 4) + (\$.50 \times 4)$$

You can see this on the graph of Charlie's budget constraint, Figure 1, below.

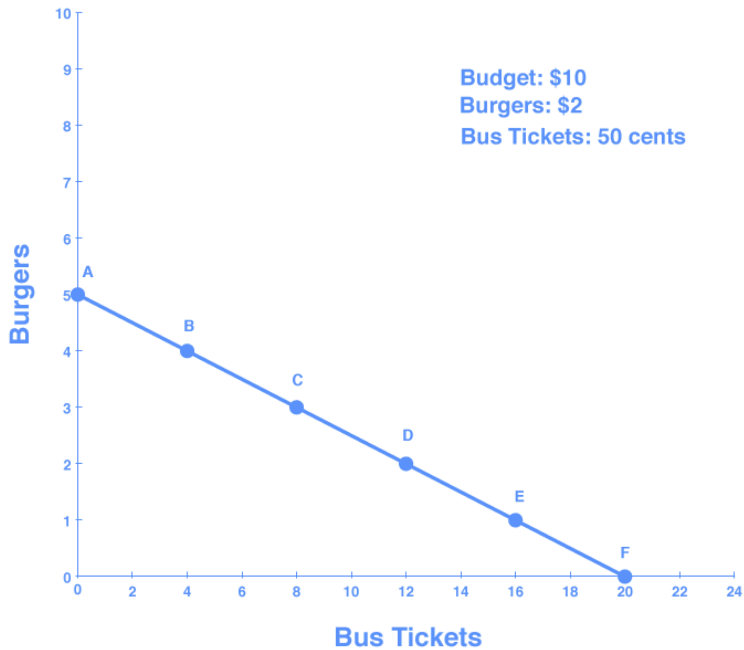


Figure 1. Charlie's Budget Constraint

If we want to answer the question "How many burgers and bus tickets can Charlie buy?" then we need to use the budget constraint equation.

Step 1. The equation for any budget constraint is the following:

$$\text{Budget } = P_1 \times Q_1 + P_2 \times Q_2 + \cdots + P_n \times Q_n$$

where P and Q are the price and respective quantity of any number, n, of items purchased and Budget is the amount of income one has to spend.

Step 2. Apply the budget constraint equation to the scenario.

In Charlie's case, this works out to be

$$\text{Budget} = P_1 \times Q_1 + P_2 \times Q_2$$
$$\text{Budget} = \$10$$
$$P_1 = \$2 \text{ (the price of a burger)}$$
$$Q_1 = \text{quantity of burgers (variable)}$$
$$P_2 = \$0.50 \text{ (the price of a bus ticket)}$$
$$Q_2 = \text{quantity of tickets (variable)}$$

For Charlie, this is

$$\$10 = \$2 \times Q_1 + \$0.50 \times Q_2$$

Step 3. Simplify the equation.

At this point we need to decide whether to solve for Q_1 or Q_2.

Remember, $Q_1 = \text{quantity of burgers}$. So, in this equation Q_1 represents the number of burgers Charlie can buy depending on how many bus tickets he wants to purchase in a given week. $Q_2 = textquantityoftickets$. So, Q_2 represents the number of bus tickets Charlie can buy depending on how many burgers he wants to purchase in a given week.

We are going solve for Q_1.

$$\begin{aligned}
10 &= 2Q_1 + 0.50Q_2 \\
10 - 2Q_1 &= 0.50Q_2 \\
-2Q_1 &= -10 + 0.50Q_2 \\
(2)\,(-2Q_1) &= (2) - 10 + (2)0.50Q_2 && \text{Clear decimal by multiplying everything by 2} \\
-4Q_1 &= -20 + Q_2 \\
Q_1 &= 5 - \tfrac{1}{4}Q_2 && \text{Divide both sides by } -4
\end{aligned}$$

Step 4. Use the equation.

Now we have an equation that helps us calculate the number of burgers Charlie can buy depending on how many bus tickets he wants to purchase in a given week.

For example, say he wants 8 bus tickets in a given week. Q_2 represents the number of bus tickets Charlie buys, so we plug in 8 for Q_2, which gives us

$$Q_1 = 5 - \left(\tfrac{1}{4}\right)8$$
$$Q_1 = 5 - 2$$
$$Q_1 = 3$$

This means Charlie can buy 3 burgers that week (point C on the graph, above).

Let's try one more. Say Charlie has a week when he walks everywhere he goes so that he can splurge on burgers. He buys 0 bus tickets that week. Q_2 represents the number of bus tickets Charlie buys, so we plug in 0 for Q_2, giving us

$$Q_1 = 5 - \left(\frac{1}{4}\right)0 Q_1 = 5$$

So, if Charlie doesn't ride the bus, he can buy 5 burgers that week (point A on the graph).

If you plug other numbers of bus tickets into the equation, you get the results shown in Table 1, below, which are the points on Charlie's budget constraint.

Table 1.

Point	**Quantity of Burgers (at $2)**	**Quantity of Bus Tickets (at 50 cents)**
A	5	0
B	4	4
C	3	8
D	2	12
E	1	16
F	0	20

Step 4. Graph the results.

If we plot each point on a graph, we can see a line that shows us the number of burgers Charlie can buy depending on how many bus tickets he wants to purchase in a given week.

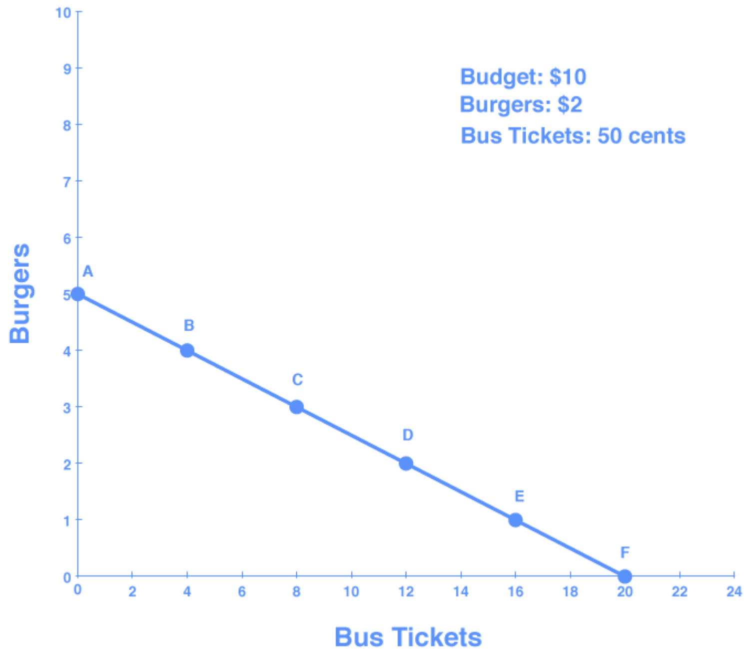


Figure 2. Charlie's Budget Constraint

We can make two important observations about this graph. First, the slope of the line is negative (the line slopes downward). Remember in the last module when we discussed graphing, we noted that when when X and Y have a negative, or inverse, relationship, X and Y move in opposite directions—that is, as one rises, the other falls. This means that the only way to get more of one good is to give up some of the other.

Second, the slope is defined as the change in the number of burgers (shown on the vertical axis) Charlie can buy for every incremental change in the number of tickets (shown on the horizontal axis) he buys. If he buys one less burger, he can buy four more bus tickets. The slope of a budget constraint always shows the opportunity cost of the good that is on the horizontal axis. If Charlie has to give up lots of burgers to buy just one bus ticket, then the slope will be steeper, because the opportunity cost is greater.

Let's look at this in action and see it on a graph. What if we change the price of the burger to $1? We will keep the price of bus tickets at 50 cents. Now, instead of buying 4 more tickets for every burger he gives up, Charlie can only buy 2 tickets for every burger he gives up. Figure 3, below, shows Charlie's new budget constraint (and the change in slope).

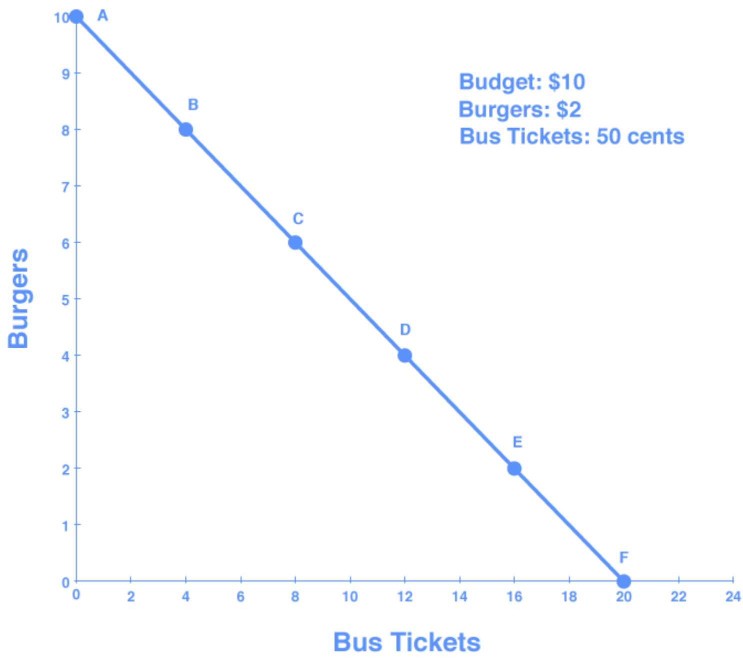


Figure 3. Charlie's New Budget Constraint

Self Check: The Cost of Choices

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does

not count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1551

30. Outcome: The Production Possibilities Frontier

What you'll learn to do: illustrate society's trade-offs by using a production possibilities frontier (or curve)

Now that we understand positive and normative statements, let's return to how individuals and societies make choices. Here we're going to focus on production and the tradeoffs involved in producing one thing versus another.

First we'll consider the example of a student, whose limited time means that doing one thing necessitates doing less of another thing. You'll learn how to show these choices and consequences graphically. A single individual making choices between two different production options is the simplest scenario, so it's important to understand this case before moving on to more complex ones.

Next, we'll explore the idea of production options on a much bigger scale—on the level of what an entire society can produce—and use what's known as the **production possibilities curve** (also called the **production possibilities frontier**) to see the different choices.

The production possibilities curve is a diagram that shows the possible combinations of two products or services that could potentially be produced within a society. For example, a country could choose to spend all of its income on defense or on education. The curve represents points where the country could maximize the use of its resources by choosing to produce a combination of defense and education, devoting all its resources to just defense, or devoting them just to education. In this section you'll gain some first-hand experience with economic models and graphing tools, too.

The specific things you'll learn in this section include the following:

- Explain the production possibilities frontier
- Identify the impact of society's production choices

Learning Activities

The learning activities for this section include the following:

- Video: Society's Production Possibilities Curve
- Reading: The Production Possibilities Frontier
- Reading: Productive Efficiency and Allocative Efficiency

31. Video: Society's Production Possibilities Curve

Now that you understand the choices made by a single individual, we'll take a look at the production possibilities for a society.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=51#oembed-1

To help you practice graphing, the data table used in the video is shown below.

	Military	Civilian
A)	5,000	0
B)	4,800	1,000
C)	4,500	2,000
D)	4,000	3,000
E)	3,300	4,000
F)	2,000	5,000
G)	0	6,000

32. Reading: The Production Possibilities Frontier

Overview

Let's review the production possibilities frontier and focus more specifically on the shape of the curve.

As a reminder, the **production possibilities frontier (PPF)** is an economic model that shows the possible combinations of two products or services that could potentially be produced by a society. Remember, an economic model is a simplified version of reality that allows us to observe, understand, and make predictions about economic behavior. With the PPF model, we're focused on a society's production choices and trade-offs.

Because society has limited resources (e.g., labor, land, capital, and raw materials) at any given moment, there's a limit to the quantities of goods and services it can produce. Suppose a society desires two products: health care and education. This situation is illustrated by the production possibilities frontier in Figure 1.

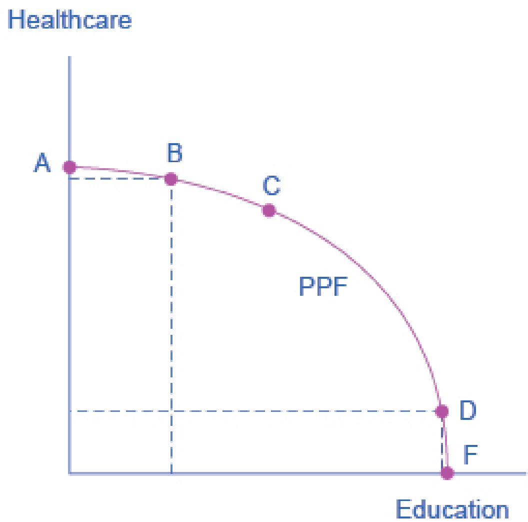


Figure 1. *Health Care vs. Education Production Possibilities Frontier.*

Figure 1 shows a trade-off between devoting resources to health care and to education. Health care is shown on the vertical axis, and education is shown on the horizontal axis. If the society were to allocate all of its resources to health care, it could produce at point A. But it would not have any resources to produce education. If it were to allocate all of its resources to education, it could produce at point F. Alternatively, the society could choose to produce any combination of health care and education shown on the production possibilities frontier.

Suppose society has chosen to operate at point B, and it's considering producing more education. Because the PPF is downward sloping from left to right, the only way society can obtain more education is by giving up some health care. That's the trade-off that society faces. Suppose it considers moving from point B to point C. What would be the opportunity cost for the additional education? The opportunity cost would be the health care that society has to give up.

Do you remember Charlie choosing combinations of burgers and bus tickets within his budget constraint? In effect, the production possibilities frontier plays the same role for society as the budget constraint plays for Charlie. Society can choose any combination of the two goods on or inside the PPF, but it doesn't have enough resources to produce outside the PPF. Just as with Charlie's budget constraint, the opportunity cost is shown by the *slope* of the production possibilities frontier.

Difference between Budget Constraint and PPF

There are differences between a budget constraint and a production possibilities frontier. A budget constraint model shows the purchase choices that an individual or society can make given a specific budget and specific purchase prices. The production possibilities frontier shows the possible combinations of two products or services that could potentially be produced by a society. Budgets and prices are more precise. If you think about it, a society's "possibilities of production" are vastly more complicated and have a great degree of variability. For this reason, a PPF is not as precise.

Consider the PPF graph above. There are no numbers on the axes of the PPF because we don't know the exact amount of resources this imaginary economy has, nor do we know how many resources it takes to produce health care and how many resources it takes to produce education. If this were a real-world example, some data would be available, but there's no single way to measure "amounts" of education and health care. That said, you could probably think of ways to measure *improvements* in education, such as more years of school completed, fewer high-school dropouts, and higher scores on standardized tests. Similarly, you could probably measure improvements in health care according to things like longer life expectancy, lower levels of infant mortality, fewer outbreaks of disease, and so on. These types of measures in a PPF are useful, but do not have the same level of accuracy as a budget constraint model.

Whether or not we have actual numbers, conceptually we can measure the opportunity cost of additional education as society moves from point B to point C on the PPF. The additional education is measured by the horizontal distance between B and C. The foregone health care is given by the vertical distance between B and C. The slope of the PPF between B and C is (approximately) the vertical distance (the "rise") over the horizontal distance (the "run"). This is the opportunity cost of the additional education.

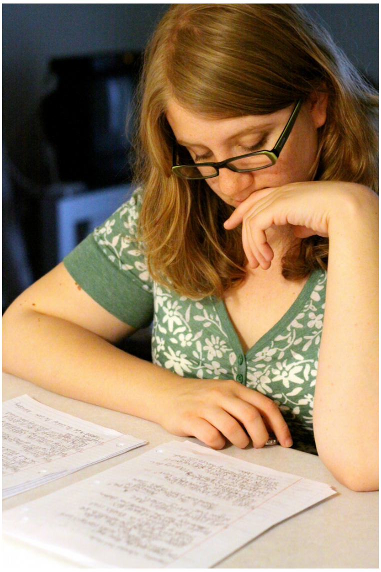
The Law of Diminishing Returns and the Curved Shape of the PPF

The budget constraints presented earlier in this module, showing individual choices about what quantities of goods to consume, were all straight lines. The reason for these straight lines was that the slope of the budget constraint was determined by the relative prices of the two goods in the *budget constraint*. However, the production possibilities frontier for health care and education was drawn as a curved line. Why does the PPF have a different shape?

To understand why the PPF is curved, start by considering point A at the top left-hand side of the PPF. At point A, all available resources are devoted to health care and none is left for education. This situation would be extreme and even ridiculous. For example, children are seeing a doctor every day, whether they're sick or not, but not attending school. People are having cosmetic surgery on every part of their bodies, but no high school or college education exists. Now imagine that some of these resources are diverted from health care to education, so that the economy is at point B instead of point A. Diverting some resources away from A to B causes relatively little reduction in health because the last few marginal dollars going into health-care services are not producing much additional gain in health. However, putting those marginal dollars into education, which is completely without resources at point A, can produce relatively large gains. For this reason, the shape of the PPF from A to B is relatively flat, representing a relatively small drop-off in health and a relatively large gain in education.

Now consider the other end, at the lower right, of the production possibilities frontier. Imagine that society starts at choice D, which is devoting nearly all resources to education and very few to health care, and it moves to point F, which is devoting *all* spending to education and none to health care. For the sake of concreteness, you can imagine that in the movement from D to F, the last few doctors must become high school science teachers, the last few nurses must become school librarians rather than dispensers of vaccinations, and the last few emergency rooms are turned into kindergartens. The gains to education from adding these last few resources to education are very small. However, the opportunity cost lost to health will be fairly large, and thus the slope of the PPF between D and F is steep, showing a large drop in health for only a small gain in education.

The lesson is not that society is likely to make an extreme choice like devoting no resources to education at point A or no resources to health at point F. Instead, the lesson is that the gains from committing additional marginal resources to education depend on how much is already being spent. If, on the one hand, very few resources are currently committed to education, then an increase in resources used can bring relatively large gains. On the other hand, if a large number of resources is already committed to education, then committing additional resources will bring relatively smaller gains.



If you've ever pulled an all-nighter, you're probably familiar with the law of diminishing returns: as the night wears on and you get tired, every additional hour you study is a little less productive than the one before.

This pattern is so common that it has been given a name: the **law of diminishing returns.** This law asserts that as additional increments of resources are devoted to a certain purpose, the marginal benefit from those additional increments will decline. For example, after not spending much at all on crime reduction, when a government spends a certain amount more, the gains in crime reduction could be relatively large. But additional increases after that typically cause relatively smaller reductions in crime, and paying for enough police and security to reduce crime to zero would be tremendously expensive.

The curve of the production possibilities frontier shows that as additional resources are added to education, moving from left to right along the horizontal axis, the initial gains are fairly large, but those gains gradually diminish. Similarly, as additional resources are added to health care, moving from bottom to top on the vertical axis, the initial gains are fairly large but again gradually diminish. In this way, the law of diminishing returns produces the outward-bending shape of the production possibilities frontier.

33. Reading: Productive Efficiency and Allocative Efficiency

Efficiency

The study of economics does not presume to tell a society what choice it should make along its production possibilities frontier. In a market-oriented economy with a democratic government, the choice will involve a mixture of decisions by individuals, firms, and government. However, economics can point out that some choices are unambiguously better than others. This observation is based on the idea of efficiency. In everyday parlance, *efficiency* refers to lack of waste. An inefficient washing machine operates at high cost, while an efficient washing machine operates at lower cost, because it's not wasting water or energy. An inefficient organization operates with long delays and high costs, while an efficient organization is focused, meets deadlines, and performs within budget.

The production possibilities frontier can illustrate two kinds of efficiency: productive efficiency and allocative efficiency. Figure 1, below, illustrates these ideas using a production possibilities frontier between health care and education.

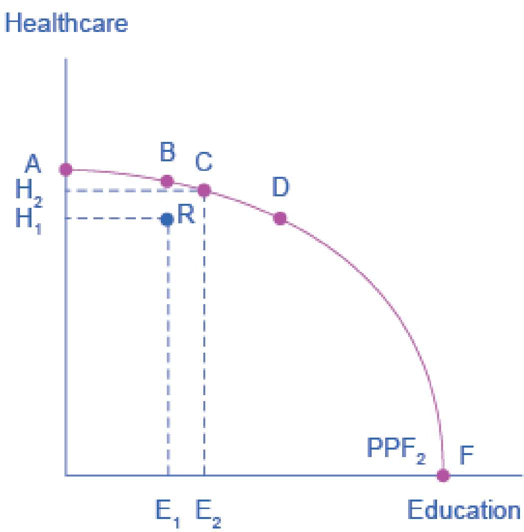


Figure 1. Productive and Allocative Efficiency.

Productive efficiency means that, given the available inputs and technology, it's impossible to produce more of one good without decreasing the quantity of another good that's produced. All choices along the PPF in Figure 1, such as points A, B, C, D, and F, display productive efficiency. As a firm moves from any one of these choices to any other, either health care increases and education decreases or vice versa. However, any choice inside the production possibilities frontier is productively inefficient and wasteful because it's possible to produce more of one good, the other good, or some combination of both goods.

For example, point R is productively inefficient because it is possible at choice C to have more of both goods: education on the horizontal axis is higher at point C than point R (E_2 is greater than E_1), and health care on the vertical axis is also higher at point C than point R (H_2 is greater than H_1).

Any time a society is producing a combination of goods that falls *along* the PPF, it is achieving productive efficiency. When the combination of goods produced falls *inside* the PPF, then the society is productively inefficient.

Allocative efficiency means that the particular mix of goods a society produces represents the combination that society most desires. For example, often a society with a younger population has a preference for production of education, over production of health care. If the society is producing the quantity or level of education that the society demands, then the society is achieving allocative efficiency. Determining "what a society desires" can be a controversial question and is often discussed in political science, sociology, and philosophy classes, as well as in economics.

At the most basic level, allocative efficiency means that producers supply the quantity of each product that consumers demand. Only one of the productively efficient choices will be the allocative efficient choice for society as a whole. For example, in order to achieve allocative efficiency, a society with a young population will invest more in education. As the population ages, the society will shift resources toward health care because the older population requires more health care than education.

In the graph (Figure 1), above, a society with a younger population might achieve allocative efficiency at point D, while a society with an older population that required more health care might achieve allocative efficiency at point B.

Why Society Must Choose

Every economy faces two situations in which it may be able to expand the consumption of all goods. In the first case, a society may discover that it has been using its resources inefficiently, in which case by improving efficiency and producing on the production possibilities frontier, it can have more of all goods (or at least more of some and less of none). In the second case, as resources grow over a period of years (e.g., more labor and more capital), the economy grows. As it does, the production possibilities frontier for a society will tend to shift outward, and society will be able to afford more of all goods.

However, improvements in productive efficiency take time to discover and implement, and economic growth happens only gradually. So, a society must choose between trade-offs in the present—as opposed to years down the road. For government, this process often involves trying to identify where additional spending could do the most good and where reductions in spending would do the least harm. At the individual and firm level, the market economy coordinates a process in which firms seek to produce goods and services in the quantity, quality, and price that people want. But for both the government and the market economy, in the short term, increases in production of one good typically mean offsetting decreases somewhere else in the economy.

The PPF and Comparative Advantage

While every society must choose how much of each good it should produce, it doesn't need to produce every single good it consumes. Often, how much of a good a country decides to produce depends on how expensive it is to produce it versus buying it from a different country. As we saw earlier, the curve of a country's PPF gives us information about the trade-off between devoting resources to producing one good versus another. In particular, its slope gives the opportunity cost of producing one more unit of the good in the x-axis in terms of the other good (in the y-axis). Countries tend to have different opportunity costs of producing a specific good, either because of different climates, geography, technology, or skills.

Suppose two countries, the U.S. and Brazil, need to decide how much they will produce of two crops: sugar cane and wheat. Due to its climate, Brazil can produce a lot of sugar cane per acre but not much wheat. Conversely, the U.S. can produce a lot of wheat per acre, but not much sugar cane. Clearly, Brazil has a lower opportunity cost of producing sugar cane (in terms of wheat) than the U.S. The reverse is also true; the U.S. has a lower opportunity cost of producing wheat than Brazil. This can be illustrated by the PPF of each country, shown in Figure 2, below.

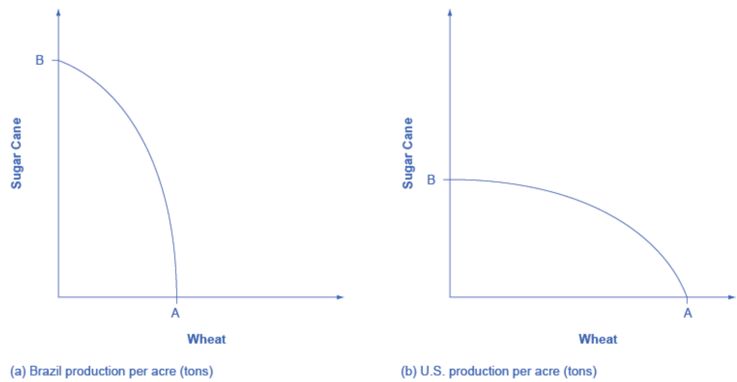


Figure 2. Brazil and U.S. PPFs

When a country can produce a good at a lower opportunity cost than another country, we say that this country has a **comparative advantage** in that good. In our example, Brazil has a comparative advantage in sugar cane, and the U.S. has a comparative advantage in wheat. One can easily see this with a simple observation of the extreme production points in the PPFs. If Brazil devoted all of its resources to producing wheat, it would be producing at point A. If, however, it devoted all of its resources to producing sugar cane instead, it would be producing a much larger amount, at point B. By moving from point A to point B, Brazil would give up a relatively small quantity in wheat production to obtain a large production in sugar cane. The opposite is true for the U.S. If the U.S. moved from point A to B and produced only sugar cane, this would result in a large opportunity cost in terms of foregone wheat production.

The slope of the PPF gives the opportunity cost of producing an additional unit of wheat. While the slope is not constant throughout the PPFs, it is quite apparent that the PPF in Brazil is much steeper than in the U.S., and therefore the opportunity cost of wheat is generally higher in Brazil. In the module on International Trade you will learn that countries' differences in comparative advantage determine which goods they will choose to produce and trade. When countries engage in trade, they specialize in the production of the goods in which they have comparative advantage and trade part of that production for goods in which they don't have comparative advantage in. With trade, goods are produced where the opportunity cost is lowest, so total production increases, benefiting both trading parties.

Self Check: The Production Possibilities Frontier

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1552

34. Outcome: Economic Rationality

What you'll learn to do: explain the assumption of economic rationality

Economists assume that human decision-making is predictable and rational. They believe that, when making choices, people basically try to avoid costs and maximize benefits to themselves. In this section, you'll learn more about the principle of economic rationality and the role it plays in economic models.

The specific things you'll learn in this section include the following:

- Define rationality in an economic context
- Provide examples of rational decision-making

Learning Activities

The learning activities for this section include the following:

- Reading: Rationality and Self-Interest
- Reading: Rationality in Action

35. Reading: Rationality and Self-Interest

If you say that someone is behaving “rationally,” you probably mean that he or she is acting in a thoughtful, clear-headed way (as opposed to *irrationally*, which suggests that someone is acting emotionally or illogically). In the context of economics, the term *rationality* has a very specific meaning. It refers to an assumption that economists make about how people behave—remember that this is the starting point of all economics—*in the face of scarcity*. There simply aren’t enough resources to satisfy all needs and wants. Charlie has only $10, he’s hungry, and he needs to get to work. What will he do? An economist predicts that Charlie will behave in a predictable, rational manner, balancing costs against benefits to arrive at an action that maximizes his personal happiness or utility:

As a result, he will choose a certain number of burgers and a certain number of bus tickets.

Economists assume that people will make choices in their own self-interest. They will choose those things that provide the greatest personal benefit, and they'll avoid or forego those that aren't as personally valuable and compelling. That's what we mean by the **assumption of rationality**.

Do economists really believe that we only think of ourselves and don't ever try to benefit others? Not at all. The assumption that individuals are purely self-interested doesn't imply that individuals are greedy and selfish. People clearly derive satisfaction from helping others, so "self-interest" can also include pursuing things that benefit other people. The assumption of rationality—also called the **theory of rational behavior**—is primarily a simplification that economists make in order to create a useful model of human decision-making.

36. Reading: Rationality in Action

If you consider your own personal choices, you will probably find that they are quite complex. You are balancing what you want right now with options you want to have in the future. You probably value the people around you—friends, family, neighbors—and you may consider the impact that your choices have on them.

Setting aside the messy realm of personal choices for the time being, let's take a look at how decisions are made by consumers and by businesses in a world of economic rationality.

Rationality and Consumers

When a consumer is thinking about buying a product, what does he or she want? The theory of rational behavior would say that the consumer wants to maximize benefit and minimize cost.

Let's look at a simple example. When a new movie is released, will you see it in the theater, or will you wait for it to be released on Netflix or on TV? If we consider only the monetary costs of your choice, a movie ticket might cost $10 and you will only be able to see that movie one time. If you wait, you can probably watch it as part of your monthly Netflix or cable subscription without spending any more than you would spend without watching the movie. Why would you pay $10 to watch the movie in the theater? You might want to see it right away, when it is only showing in the theater. You might want the theater experience, with the big screen and high-quality image and sound. You will make a decision that is economically rational, based on the following consideration: "Is the benefit and enjoyment that I get from seeing the movie in a theater worth the $10 cost?"

As a consumer, you are making an economically rational decision about the costs and benefits.

Since we will build upon this later in the course, it's important to understand that this assumption creates a link between the cost of a product and the degree to which a consumer will want to buy it. As the cost of the product increases, it becomes less likely that the consumer will decide that the benefits of the purchase outweigh the costs.

Rationality and Businesses

Businesses also have predictable behavior, but rather than seeking to maximize happiness or pleasure, they seek to maximize profits.

When economists assume that businesses have a goal of maximizing profits, they can make predictions about how companies will react to changing business conditions.

For example, if wages in the United States increase, how will U.S. companies react? The rational reaction may be to move those jobs that can be performed elsewhere to countries with lower wages. This prediction is based on an oversimplification, and it might not hold true in every case—individual businesses would obviously need to understand the full cost of moving certain work out of the country before doing so. But the decision would be made according to the impact on profit and would still be rational. If a company stands to earn more profit by moving some jobs overseas, then that's the result that economists would predict.

Rationality suggests that consumers will act to maximize self-interest and businesses will act to maximize profits. Both are taking into account the benefits of a choice, given the costs.

Self Check: Economic Rationality

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1553

37. Outcome: Marginal Analysis

What you'll learn to do: define marginal analysis

Economists recognize that very few choices in the real world are "all or nothing." Most of the time, people have the choice to do a little more or a little less of something: Should you eat one more muffin? Should you study economics for another hour? Should you spend a little less money on gas? Economists use the word *marginal* to mean "additional" or "extra," and they use the term *marginal analysis* to describe how people make choices by comparing the benefits and costs of doing a bit more or a bit less.

The specific things you'll learn in this section include the following:

- Define marginal cost
- Define marginal benefit

Learning Activities

The learning activities for this section include the following:

- Reading: Marginal Analysis

38. Reading: Marginal Analysis

A Little More or a Little Less

The budget constraint framework helps to illustrate that most choices in the real world are not about getting all of one thing or all of another—we rarely decide "all burgers" or "all bus tickets." Options usually fall somewhere on a continuum, and the choice usually involves marginal decision-making and marginal analysis.

Marginal decision-making means considering a little more or a little less than what we already have. We decide by using **marginal analysis**, which means comparing the costs and benefits of a little more or a little less.

It's natural for people to compare costs and benefits, but often we look at total costs and total benefits, when the best choice requires comparing how costs and benefits change from one option to another. In short, you might think of marginal analysis as "change analysis." Marginal analysis is used throughout economics. This subtle concept is easier to grasp with examples.

Marginal Cost

Generally speaking, **marginal cost** is the difference (or change) in cost of a different choice. From a consumer's point of view, marginal cost is the additional cost of one more item purchased. From a business's point of view, marginal cost is the additional cost of one more item produced.

Suppose you typically spend a week at the beach for vacation, but this year you earned an annual bonus from your job. Should you rent a beach house for one week or two? A one-week rental costs $2,000. A two-week rental costs $3,600. Holding everything else constant, which option is better? If you stay for two weeks, the cost is significantly higher: $3,600 versus $2,000. But consider the cost by week. The first week costs $2,000. The difference in cost between one week and two is $3,600 – $2,000, or $1,600. Thus, while the marginal cost of the first week's rental is $2,000, the marginal cost of the second week's rental is $1,600. This illustrates the key rule of marginal analysis: Marginal cost = the change in total cost from one option to another.

Consider another example. Imagine that you're out getting ice cream with your friends or family. You can choose whether to buy one, two, or three scoops of ice cream. One scoop costs $3.00, two scoops cost $5.00, and three scoops cost $7.00. This information is shown in the following table.

Scoops of Ice Cream	1	2	3
Total Cost	$3	$5	$7

What is the marginal cost of each scoop of ice cream? The marginal cost of the first scoop of ice cream is $3.00 because you have to pay $3.00 more to get one scoop of ice cream than you do to get zero scoops of ice cream. The marginal cost of the second scoop of ice cream is $2.00 because you only need to pay two more dollars to get two scoops than you need to pay to get one scoop. The marginal cost of the third scoop is also $2.00 because you would need to pay an additional two dollars to get that third scoop.

Scoops of Ice Cream	1	2	3
Marginal Cost	$3	$2	$2

Marginal costs sometimes go up and sometimes go down, but to get the clearest view of your options, you should always try to make decisions based on marginal costs, rather than total costs.

Marginal Benefit

Generally speaking, **marginal benefit** is the difference (or change) in what you receive from a different choice. From a consumer's point of view, marginal benefit is the additional satisfaction of one more item purchased. From a business' point of view, marginal benefit is the additional revenues received from selling one more item.

Suppose you're considering membership at the local recreation center. The basic membership gives access to the swimming pool, while the full membership gives access to the swimming pool and the weight room. What is the difference between the two memberships? Since both give access to the pool, the marginal benefit of full membership is access to the weight room.

The amount of benefit a person receives from a particular good or service is subjective; one person may get more satisfaction or happiness from a particular good or service than another. For example, you might enjoy ice cream more than your friend who is allergic to dairy. The amount of benefit you get can also change. For example, you might enjoy the ice cream more on a hot day than on a cold day. This doesn't make it any less real, however.

Economic Rationality Revisited

How, then, do you decide on a choice? The answer is that you compare, to the best of your ability, the marginal benefits with the marginal costs. An economically rational decision is one in which the marginal benefits of a choice are greater than the marginal costs of the choice.

If we return to the recreation center example above, suppose that the basic membership is $30 per month, while the full membership is $40 per month. An economically rational decision-maker would ask, Is the marginal benefit (access to the weight room) worth the marginal cost (an extra $10 per month)? For some people, the answer will be yes. For others, it will be no. Either way, marginal analysis is an important part of economic rationality and good decision-making.

Self Check: Marginal Analysis

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1554

39. Outcome: Positive and Normative Statements

What you'll learn to do: differentiate between positive and normative statements

What choices does society have about how it uses its resources?—"What is possible?" we might ask. Also, what is the right thing to do with the resources we have?—"What *should* we choose," we might ask. These are very different questions that lead to very different kinds of responses and statements.

In this section we are going to consider the difference between positive and normative statements and their role in economics. **Positive statements** are objective. **Normative statements** are subjective. Good economists are careful to differentiate between the two. In this section, we will learn to differentiate between descriptions of the world as it is and the world as it should be.

The specific things you'll learn in this section include the following:

- Define positive and normative statements
- Provide examples of positive and normative statements

Learning Activities

The learning activities for this section include the following:

- Video: Positive and Normative Analysis

- Reading: Positive and Normative Statements

40. Video: Positive and Normative Analysis

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=60#oembed-1

41. Reading: Positive and Normative Statements

Economics seeks to describe economic behavior as it actually exists, and it relies on a distinction between **positive statements**, which describe the world as it is, and **normative statements**, which describe how the world should be.

Positive Statements

Two kinds of assertions in economics can be subjected to testing. One is the hypothesis. Another testable assertion is a statement of fact, such as “It’s raining,” or “Microsoft is the largest producer of computer operating systems in the world.” Like hypotheses, such assertions can be shown to be correct or incorrect. A statement of fact or a hypothesis is a positive statement.

Normative Statements

Although people often disagree about positive statements, such disagreements can ultimately be resolved through investigation. There is another category of assertions, however, for which investigation can never resolve differences. A normative statement is one that makes a value judgment. Such a judgment is the opinion of the speaker; no one can "prove" that the statement is or is not correct. Here are some examples of normative statements in economics:

- We ought to do more to help the poor.
- People in the United States should save more for retirement.
- Corporate profits are too high.

These statements are based on the values of the person who makes them and can't be proven false.

Because people have different values, normative statements often provoke disagreement. An economist whose values lead him or her to conclude that we should provide more help for the poor will disagree with one whose values lead to a conclusion that we should not. Because no test exists for these values, these two economists will continue to disagree, unless one persuades the other to adopt a different set of values. Many of the disagreements among economists are based on such differences in values and therefore are unlikely to be resolved.

Positive and Normative Analysis Video

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=61#oembed-1

Self Check: Positive and Normative Statements

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1555

42. Problem Set: Budget Constraints and Opportunity Cost

Test your understanding of the learning outcomes in this module by working through the following problems. These problems aren't graded, but they give you a chance to practice before taking the quiz.

If you'd like to try a problem again, you can click the link that reads, "Try another version of these questions."

Use the information provided in the first question for all of the questions in this problem set.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=62

43. Problem Set: Marginal Analysis

Test your understanding of the learning outcomes in this module by working through the following problems. These problems aren't graded, but they give you a chance to practice before taking the quiz.

If you'd like to try a problem again, you can click the link that reads, "Try another version of this question."

Farmer's Market

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=63

Midterm Grades

One or more interactive elements has been excluded from this version of the text. You can view them online *here: https://library.achievingthedream.org/sacmicroeconomics/?p=63*

44. Putting It Together: Choice in a World of Scarcity

Summary

In this module you learned that the study of economics is about how choices are made by individuals and entities, given the fact that we can never have enough of the things we want. You learned how to:

- Explain the cost of choices and trade-offs
- Illustrate society's trade-offs by using a production possibilities frontier (or curve)
- Explain the assumption of economic rationality by individuals and firms
- Define marginal analysis
- Differentiate between positive and normative statements

The Challenging Budget Constraints of a Student

We began this module with a discussion of the annual salaries of full-time U.S. workers with different levels of education. Let's return to the very real economic issues that face most students when making decisions about their education.

First, we discussed the cost of choices and trade-offs and used the budget constraint model to demonstrate those costs. Each term, students make a trade-off between taking more credits in school and buying necessary items. Let's create a budget constraint model for Camila, a community college student who is struggling to cover the cost of education. First, let's assume that each credit hour costs

$75. Camila wants to take 12 to 16 credits but also needs to pay for gas to drive between school, work, and other family responsibilities. Gas costs $3 per gallon. If she has a budget during the course of the academic term that allows her to spend a total of $1,500 on course credits and gas, what are Camila's options?

We can use the budget constraint equation to answer this question.

Step 1. Apply the budget constraint equation to the scenario.

In Camila's case, this works out to be

$$\text{Budget} = P_1 \times Q_1 + P_2 \times Q_2$$
$$\text{Budget} = 1500$$
$$P_1 = 3\,(\text{price for a gallon of gas})$$
$$Q_1 = \text{gallons of gas (variable)}$$
$$P_2 = 75\,(\text{price per credit hour})$$
$$Q_2 = \text{number of credit hours (variable)}$$

For Camila, this is

$$1500 = 3 \times Q_1 + 75 \times Q_2$$

Step 2. Simplify the equation.

At this point we need to decide whether to solve for Q_1 or Q_2.

Remember, Camila was hoping to take at least 12 credit hours, so we know the value for Q_2. We will solve for Q_1 because, in this equation, it represents the number of gallons of gas Camila can pay for, depending on how many credit hours she takes during the academic term.

We are going solve for Q_1. First we will write the equation with the variables on the left to make solving easier: $3Q_1 + 75Q_2 = 1500$.

\begin{array}{lll}3Q_1+75Q_2=1500\\\\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,3Q_1=1500-75Q_2\,\,\,\,\,\,\,\,\,\,\,\,\text{isolate Q_1 on one side}\\\\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\large\frac{3Q_1}{3}=\frac{1500}{3}-\frac{-75Q_2}{3}\,\,\,\,\,\,\,\,\,\normalsize\text{divide everything by 3}\\\\,Q_1=500-25Q_2\end{array}

Step 3. Use the equation.

We know that Camila hopes to take 12 credit hours during a term. Q_1 represents the number of credits she hopes to fund, so we plug in 12 for Q_2, which gives us

$$Q_1 = 500 - 25(12)$$
$$Q_1 = 500 - 300$$
$$Q_1 = 200$$

This means that Camila can buy 200 gallons of gas during the term she is taking 12 credit hours (point M on the graph, below).

If you plug other numbers of credit hours into the equation, you get the results shown in Table 1, below.

Table 1. Camila's Budget Constraint		
Point	**Number of Credit Hours**	**Gallons of Gas**
A	0	500
B	1	475
C	2	450
D	3	425
E	4	400
F	5	375
G	6	350
H	7	325
I	8	300
J	9	275
K	10	250
L	11	225
M	12	200
N	13	175
O	14	150
P	15	125
Q	16	100

Step 4. Graph the results.

If we plot each point on a graph, as below, we can see a line that shows us the number of credit hours that Camila can fund while still paying for gas.

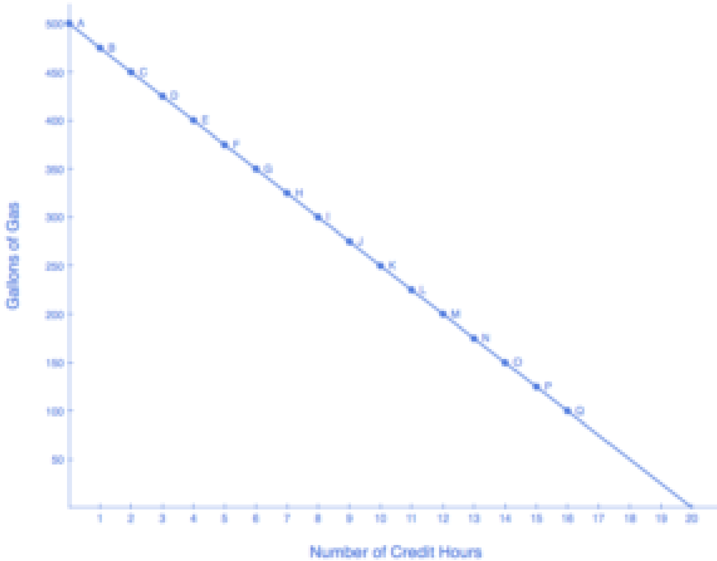


Figure 1. Camila's Budget Constraint

Education and the Production Possibilities Curve

As state legislators allocate funding, they often make independent decision about the funding amount and approach for education and the funding amount and approach for corrections (or prisons). Economists recognize that these are not independent decisions. The production possibilities curve demonstrates that if society invests more in prisons, there are will be a reduction in the resources available to invest in education.

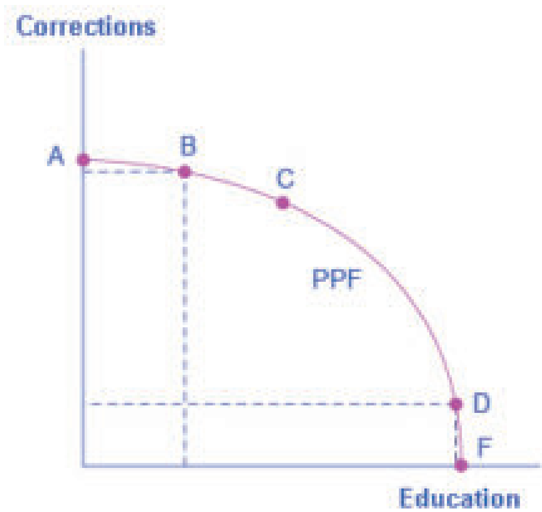


Figure 2. Education vs. Corrections Production Possibilities Frontier

The graph above demonstrates the trade-off between devoting resources to corrections and to education. If the society were to allocate all of its resources to corrections, it could produce at point A, but it would not have any resources to produce education. If it were to allocate all of its resources to education, it could produce at point F. Alternatively, society could choose to produce any combination of corrections and education shown on the production possibilities frontier.

Sometimes legislators don't recognize the direct trade-off between investing in education and investing in prisons, but inevitably economists will point out the connection, and the press will jump in and question the legislators' decision. In a world of scarcity, more spending in one necessarily means less to spend in others.

Should Society Invest in Prisons or Education?

As we consider the trade-offs between investments in prisons and education, is there a definitive "right" answer? Consider the following analysis by the Center on Budget and Policy Priorities:

> Even as states spend more on corrections, they are underinvesting in educating children and young adults, especially those in high-poverty neighborhoods. At least 30 states are providing less general funding per student this year for K–12 schools than before the recession, after adjusting for inflation; in 14 states the reduction exceeds 10 percent. Higher education cuts have been even deeper: the average state has cut higher education funding per student by 23 percent since the recession hit, after adjusting for inflation. Eleven states spent more of their general funds on corrections than on higher education in 2013. And some of the states with the biggest education cuts in recent years also have among the nation's highest incarceration rates.
>
> This is not sound policy. State economies would be much stronger over time if states invested more in education and other areas that can boost long-term economic growth and less in maintaining extremely high prison populations. The economic health of many low-income neighborhoods, which face disproportionately high incarceration rates, could particularly improve if states reordered their spending in such a way. States could use the freed-up funds in a number of ways, such as expanding access to high-quality preschool, reducing class sizes in high-poverty schools, and revising

state funding formulas to invest more in high-poverty neighborhoods.[1]

While the analysis cited is thorough and logical, the report above includes a range of positive and normative statements. If you reread the analysis with that in mind, you will find examples of both.

Positive Statements

- At least 30 states are providing less general funding per student this year for K–12 schools than before the recession, after adjusting for inflation; in 14 states the reduction exceeds 10 percent.
- Higher education cuts have been even deeper: the average state has cut higher education funding per student by 23 percent since the recession hit, after adjusting for inflation.
- Eleven states spent more of their general funds on corrections than on higher education in 2013. And some of the states with the biggest education cuts in recent years also have among the nation's highest incarceration rates.

Normative Statements

- Even as states spend more on corrections, they are underinvesting in educating children and young adults, especially those in high-poverty neighborhoods.

1. http://www.cbpp.org/sites/default/files/atoms/files/10-28-14sfp.pdf

- This is not sound policy.
- State economies would be much stronger over time if states invested more in education and other areas that can boost long-term economic growth and less in maintaining extremely high prison populations.

As you can see, your experience as a student affords you an important view into the trade-offs that are core to economics.

45. Glossary: Choice in a World of Scarcity

allocative efficiency when the mix of goods being produced represents the mix that society most desires

assumption of rationality the assumption that people will make choices in their own self-interest, choosing things that provide the greatest personal benefit and foregoing those that aren't as personally valuable and compelling; *also called the theory of rational behavior*

budget constraint all possible combinations of goods that someone can afford, given the prices of goods, when all income (or time) is spent

comparative advantage the ability of a group or country to produce a good or service at a lower opportunity cost than another group or country

law of diminishing returns as additional increments of resources are devoted to a certain purpose, the marginal benefit from those additional increments will decline

marginal analysis comparing the costs and benefits of a little more or a little less

marginal benefit is the difference (or change) in what you receive from a different choice

marginal cost the difference (or change) in cost of a different choice

normative statements are subjective; they describe the world as it ought to be

opportunity cost is the value of the next best alternative

positive statements are objective; they describe the world as it is

production possibilities frontier (or curve) a diagram that shows the productively efficient combinations of two products that an economy can produce given the resources it has available

productive efficiency when it's impossible to produce more of one good (or service) without decreasing the quantity produced of another good (or service)

sunk costs costs that are incurred in the past and can't be recovered

46. Discussion: Is Economics a Science?

Is economics a science? Why, or why not? As part of your response and explanation, include the definitions of "science" and "economics" as you understand them.

PART IV

MODULE: SUPPLY AND DEMAND

47. Why It Matters: Supply and Demand

Why analyze how buyers and sellers interact in a free and competitive market to determine prices and quantities of goods?

Do you pay attention to the cost of a cup of coffee? Most people recognize that when they make coffee at home it's cheaper than buying a cup of coffee that someone else has made. You've probably also noticed that some coffee places are more expensive than others—a cup of coffee at Starbucks usually costs more than one at a gas station, for instance. Regardless of where you decide to buy coffee, the price can change dramatically.

Figure 1. Coffee Prices. *Source: Trading Economics*

While retailers make decisions about how much they will mark up the coffee drinks they sell, the underlying coffee prices all around the world are driven by supply and demand. Brazil accounts for 33 percent of all coffee production in the world, and in 2011 the country experienced a drought. Coffee trees had already been weakened by environmental factors, and the drought had a significant impact on production levels.

How did individuals react to this kind of shortage? Did the shortage have an impact on price? Take a look at the graph in Figure 1, above. You can see from the sharp spike in 2011 that a shortage in the supply of coffee did indeed have an impact on price. But what do we know about the demand for coffee during that time? Did coffee consumption levels fall? To answer these questions, we need to know more about how buyers and sellers interact in the marketplace. In short, we need to understand supply and demand.

In this section you'll learn about these key economic factors and the laws that govern them. Understanding supply and demand is not only essential to the study of economics—it may also help you be a better-informed consumer and make knowledgeable decisions about everything from your next cup of Joe to your next job.

Learning Outcomes

- Describe and differentiate between the major economic systems
- Explain the determinants of demand
- Explain the determinants of supply
- Define and graphically illustrate market equilibrium, surplus, and shortage

48. Outcome: Economic Systems

What you'll learn to do: describe and differentiate between major economic systems

Think about what a complex system a modern economy is. It includes all production of goods and services, all buying and selling, all employment. The economic life of every individual is interrelated, at least to a small extent, with the economic lives of thousands or even millions of other individuals. Who organizes and coordinates this system? Who insures that, for example, the number of televisions a society produces is the same as the amount it needs and wants? Who insures that the right number of employees works in the electronics industry? Who insures that televisions are produced in the best way possible? How does it all get done?

The answer to these important questions depends on the kind of economic system a society uses.

In this section, you'll learn about the basic organizing principles of different types of economies. Understanding the characteristics of a competitive market, in particular, is an important foundation for understanding the mechanisms of supply and demand.

The specific things you'll learn in this section include the following:

- Define and give an example of a market economy
- Define and give an example of a planned economy
- Define and give an example of a command economy
- Define and explain the characteristics of a competitive market

Learning Activities

The learning activities for this section include the following:

- Reading: Economic Systems

49. Reading: Economic Systems

Legoland, Billund, Denmark: Picture of a Planned Economy?

Types of Economies

In the modern world today, there is a range of economic systems, from market economies to planned (or command) economies.

Market Economies

A **market** is any situation that brings together buyers and sellers of goods or services. Buyers and sellers can be either individuals or businesses. In a **market economy**, economic decision-making happens through markets. Market economies are based on private enterprise: the means of production (resources and businesses) are owned and operated by private individuals or groups of private individuals. Businesses supply goods and services based on demand. Which goods and services are supplied depends on what is demanded by consumers or other businesses. A person's income is based on his or her ability to convert resources (especially labor) into something that society values. The more society values the person's output, the higher the income they will earn (think Lady Gaga or LeBron James).

Examples of free-market economies include Hong Kong, Singapore, Australia, and the United States.

Free Markets

In a **market economy**, decisions about what products are available and at what prices are determined through the interaction of supply and demand. A **competitive market** is one in which there is a large number of buyers and sellers, so that no one can control the market price. A **free market** is one in which the government does not intervene in any way. A free and competitive market economy is the ideal type of market economy, because what is supplied is exactly what consumers demand.

Price controls are an example of a market that is not free. When government intervenes, the market outcomes will be different from those that would occur in a free and competitive market model. When markets are less than perfectly competitive (e.g., monopolistic), the market outcomes will also differ.

Planned (or Command) Economies

Command economies operate very differently. In a **command economy**, economic effort is devoted to goals passed down from a ruler or ruling class. Ancient Egypt was a good example: A large part of economic life was devoted to building pyramids (like the one at the left), for the pharaohs. Medieval manor life is another example: The lord provided the land for growing crops and protection in the event of war. In return, vassals provided labor and soldiers to do the lord's bidding. In the last century, communism emphasized command economies.

In a command economy, resources and businesses are owned by the government. The government decides what goods and services will be produced and what prices will be charged for them. The government decides what methods of production will be used and how much workers will be paid. Some necessities like health care and education are provided for free, as long as the state determines that you need them. Currently, North Korea and Cuba have command economies.

The primary distinction between a free and command economy is the degree to which the government determines what can be produced and what prices will be charged. In a free market, these determinations are made by the collective decisions of the market itself (which is comprised of producers and consumers). Producers and consumers make rational decisions about what will satisfy their self-interest and maximize profits, and the market responds accordingly. In a planned economy, the government makes most decisions about what will be produced and what the prices will be, and the market must follow that plan.

Most economies in the real world are mixed; they combine elements of command and market systems. The U.S. economy is positioned toward the market-oriented end of the spectrum. Many countries in Europe and Latin America, while primarily market-oriented, have a greater degree of government involvement in economic decisions than in the U.S. economy. China and Russia, while they are closer now to having a market-oriented system than several decades ago, remain closer to the command-economy end of the spectrum.

The following Crash Course video provides additional information about the broad economic choices that countries make when they decide between planned and market economies. The narrators talk fast, so you'll need to listen closely and possibly watch the video a second time!

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=70#oembed-1

Economic systems determine the following:

- What to produce?
- Who to produce it?
- Who gets it?

In a planned economy, government controls the factors of production:

- In a true communist economy, there is no private property—everyone owns the factors of production. This type

of planned economy is called a *command economy*
- In a socialist economy, there is some private property and some private control of industry.

In a free-market (capitalist) economy, individuals own the factors of production:

- Businesses produce products.
- Consumers choose the products they prefer causing the companies that product them to make more profit.

Even in free markets, governments will

- Maintain the rule of law
- Create public goods and services such as roads and education
- Step in when the market gets things wrong (e.g., setting minimum wage, establishing environmental standards)

In reality, economies are neither completely free-market nor completely planned. Neither exists in "pure" form, since all societies and governments regulate their economies to varying degrees. Throughout this course we will consider a number of ways in which the U.S. government influences and controls the economy.

Self Check: Economic Systems

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether

to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1558

50. Outcome: Demand

What you'll learn to do: explain the determinants of demand

Imagine that the price of Ben & Jerry's ice cream decreases by 25 percent during the next summer. What do you think will happen to the amount of Ben & Jerry's ice cream that people will want to buy? Clearly, the demand for ice cream will increase. By the same token, if the price of the ice cream were to rise by 25 percent, then the demand for the ice cream would fall. In this section, you will examine the law of demand and see why this simple concept is essential to understanding economics.

The specific things you'll learn in this section include the following:

- Explain the law of demand
- Explain a demand curve
- Create a demand curve using a data set
- Describe the differences between changes in demand and changes in the quantity demanded
- Explain the impact of factors that change demand

Learning Activities

The learning activities for this section include the following:

- Reading: What Is Demand?
- Video: Change in Demand vs. Quantity Demanded
- Reading: Factors Affecting Demand

- Worked Example: Shift in Demand
- Reading: Summary of Factors That Change Demand
- Simulation: Demand for Food Trucks
- Self Check: Demand

51. Reading: What Is Demand?

Demand for Goods and Services

Economists use the term **demand** to refer to the amount of some good or service consumers are willing and able to purchase at each price. Demand is based on needs and wants—a consumer may be able to differentiate between a need and a want, but from an economist's perspective, they are the same thing. Demand is also based on ability to pay. If you can't pay for it, you have no effective demand.

What a buyer pays for a unit of the specific good or service is called the **price**. The total number of units purchased at that price is called the **quantity demanded**. A rise in the price of a good or service almost always decreases the quantity of that good or service demanded. Conversely, a fall in price will increase the quantity demanded. When the price of a gallon of gasoline goes up, for example, people look for ways to reduce their consumption by combining several errands, commuting by carpool or mass transit, or taking weekend or vacation trips closer to home. Economists call this inverse relationship between price and quantity demanded the **law of demand**. The law of demand assumes that all other variables that affect demand are held constant.

An example from the market for gasoline can be shown in the form of a table or a graph. (Refer back to "Reading: Creating and Interpreting Graphs" in module 0 if you need a refresher on graphs.) A table that shows the quantity demanded at each price, such as Table 1, is called a **demand schedule**. Price in this case is measured in dollars per gallon of gasoline. The quantity demanded is measured in millions of gallons over some time period (for example, per day or per year) and over some geographic area (like a state or a country).

Table 1. Price and Quantity Demanded of Gasoline

Price (per gallon)	Quantity Demanded (millions of gallons)
$1.00	800
$1.20	700
$1.40	600
$1.60	550
$1.80	500
$2.00	460
$2.20	420

A **demand curve** shows the relationship between price and quantity demanded on a graph like Figure 1, below, with quantity on the horizontal axis and the price per gallon on the vertical axis. Note *that this is an exception to the normal rule in mathematics that the independent variable* (x) *goes on the horizontal axis and the dependent variable* (y) *goes on the vertical.* Economics is *different from math!*

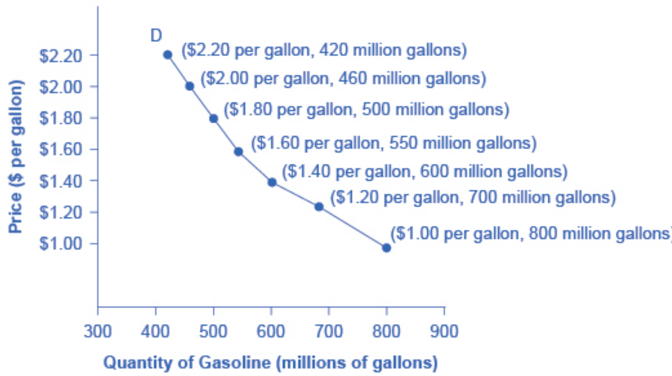


Figure 1. A Demand Curve for Gasoline

The demand schedule (Table 1) shows that as price rises, quantity demanded decreases, and vice versa. These points can then be graphed, and the line connecting them is the demand curve (shown by line D in the graph, above). The downward slope of the demand curve again illustrates the law of demand—the inverse relationship between prices and quantity demanded.

The demand schedule shown by Table 1 and the demand curve shown by the graph in Figure 1 are two ways of describing the same relationship between price and quantity demanded.

Demand curves will look somewhat different for each product. They may appear relatively steep or flat, or they may be straight or curved. Nearly all demand curves share the fundamental similarity that they slope down from left to right. In this way, demand curves embody the law of demand: As the price increases, the quantity demanded decreases, and conversely, as the price decreases, the quantity demanded increases.

Demand vs. Quantity Demanded

In economic terminology, *demand* is not the same as *quantity demanded*. When economists talk about demand, they mean the relationship between a range of prices and the quantities demanded at those prices, as illustrated by a demand curve or a demand schedule. When economists talk about quantity demanded, they mean only a certain point on the demand curve, or one quantity on the demand schedule. In short, demand refers to the curve and quantity demanded refers to the (specific) point on the curve.

The Law of Demand Video

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=72#oembed-1

The law of demand states that, other things being equal,

- More of a good will be bought the lower its price
- Less of a good will be bought the higher its price

Ceteris paribus means "other things being equal."

52. Video: Change in Demand vs. Change in Quantity Demanded

https://youtu.be/aTSwcXJ700c

A change in price does not move the demand curve. It only shows a difference in the quantity demanded.

The demand curve will move left or right when there is an underlying change in demand at all prices.

53. Reading: Factors Affecting Demand

Introduction

We defined demand as the amount of some product that a consumer is *willing* and *able* to purchase at each *price*. This suggests at least two factors, in addition to price, that affect demand. "Willingness to purchase" suggests a desire to buy, and it depends on what economists call tastes and preferences. If you neither need nor want something, you won't be willing to buy it. "Ability to purchase" suggests that income is important. Professors are usually able to afford better housing and transportation than students, because they have more income. The prices of related goods can also affect demand. If you need a new car, for example, the price of a Honda may affect your demand for a Ford. Finally, the size or composition of the population can affect demand. The more children a family has, the greater their demand for clothing. The more driving-age children a family has, the greater their demand for car insurance and the less for diapers and baby formula.

These factors matter both for demand by an individual and demand by the market as a whole. Exactly how do these various factors affect demand, and how do we show the effects graphically? To answer those questions, we need the *ceteris paribus* assumption.

The *Ceteris Paribus* Assumption

A *demand curve* or a *supply curve* (which we'll cover later in this module) is a relationship between two, and only two, variables: quantity on the horizontal axis and price on the vertical axis. The assumption behind a demand curve or a supply curve is that no relevant economic factors, other than the product's price, are changing. Economists call this assumption ***ceteris paribus***, a Latin phrase meaning "other things being equal." Any given demand or supply curve is based on the *ceteris paribus* assumption that all else is held equal. (You'll recall that economists use the *ceteris paribus* assumption to simplify the focus of analysis.) Therefore, a demand curve or a supply curve is a relationship between two, and only two, variables *when all other variables are held equal*. If all else is not held equal, then the laws of supply and demand will not necessarily hold.

Ceteris paribus is typically applied when we look at how changes in price affect demand or supply, but *ceteris paribus* can also be applied more generally. In the real world, demand and supply depend on more factors than just price. For example, a consumer's demand depends on income, and a producer's supply depends on the cost of producing the product. How can we analyze the effect on demand or supply if multiple factors are changing at the same time—say price rises and income falls? The answer is that we examine the changes one at a time, and assume that the other factors are held constant.

For example, we can say that an increase in the price reduces the amount consumers will buy (assuming income, and anything else that affects demand, is unchanged). Additionally, a decrease in income reduces the amount consumers can afford to buy (assuming price, and anything else that affects demand, is unchanged). This is what the *ceteris paribus* assumption really means. In this particular case, after we analyze each factor separately, we can combine the results. The amount consumers buy falls for two reasons: first because of the higher price and second because of the lower income.

The Effect of Income on Demand

Let's use income as an example of how factors other than price affect demand. Figure 1 shows the initial demand for automobiles as D_0. At point Q, for example, if the price is $20,000 per car, the quantity of cars demanded is 18 million. D_0 also shows how the quantity of cars demanded would change as a result of a higher or lower price. For example, if the price of a car rose to $22,000, the quantity demanded would decrease to 17 million, at point R.

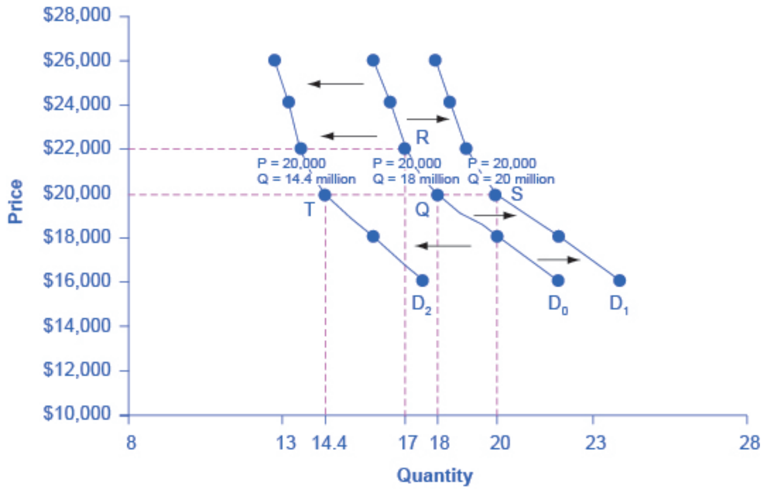


Figure 1. Shifts in Demand: A Car Example

The original demand curve D_0, like every demand curve, is based on the *ceteris paribus* assumption that no other economically relevant factors change. Now imagine that the economy expands in a way that raises the incomes of many people, making cars more affordable. How will this affect demand? How can we show this graphically?

Return to Figure 1. The price of cars is still $20,000, but with higher incomes, the quantity demanded has now increased to 20 million cars, shown at point S. As a result of the higher income levels, the demand curve shifts to the right to the new demand curve D_1, indicating an increase in demand. Table 1, below, shows clearly that this increased demand would occur at every price, not just the original one.

Table 1. Price and Demand Shifts: A Car Example

Price	Decrease to D_2	Original Quantity Demanded D_0	Increase to D_1
$16,000	17.6 million	22.0 million	24.0 million
$18,000	16.0 million	20.0 million	22.0 million
$20,000	14.4 million	18.0 million	20.0 million
$22,000	13.6 million	17.0 million	19.0 million
$24,000	13.2 million	16.5 million	18.5 million
$26,000	12.8 million	16.0 million	18.0 million

Now, imagine that the economy slows down so that many people lose their jobs or work fewer hours, reducing their incomes. In this case, the decrease in income would lead to a lower quantity of cars demanded at every given price, and the original demand curve D_0 would shift left to D_2. The shift from D_0 to D_2 represents such a decrease in demand: At any given price level, the quantity demanded is now lower. In this example, a price of $20,000 means 18 million cars sold along the original demand curve, but only 14.4 million sold after demand fell.

When a demand curve shifts, it does not mean that the quantity demanded by every individual buyer changes by the same amount. In this example, not everyone would have higher or lower income and not everyone would buy or not buy an additional car. Instead, a shift in a demand curve captures a pattern for the market as a whole: Increased demand means that at every given price, the quantity demanded is higher, so that the demand curve shifts to the right from D_0 to D_1. And, decreased demand means that at every given price, the quantity demanded is lower, so that the demand curve shifts to the left from D_0 to D_2.

We just argued that higher income causes greater demand at every price. This is true for most goods and services. For some—luxury cars, vacations in Europe, and fine jewelry—the effect of a rise in income can be especially pronounced. A product whose demand rises when income rises, and vice versa, is called a **normal**

good. A few exceptions to this pattern do exist, however. As incomes rise, many people will buy fewer generic-brand groceries and more name-brand groceries. They are less likely to buy used cars and more likely to buy new cars. They will be less likely to rent an apartment and more likely to own a home, and so on. A product whose demand falls when income rises, and vice versa, is called an **inferior good**. In other words, when income increases, the demand curve shifts to the left.

Other Factors That Shift Demand Curves

Income is not the only factor that causes a shift in demand. Other things that change demand include tastes and preferences, the composition or size of the population, the prices of related goods, and even expectations. A change in any one of the underlying factors that determine what quantity people are willing to buy at a given price will cause a shift in demand. Graphically, the new demand curve lies either to the right (an increase) or to the left (a decrease) of the original demand curve. Let's look at these factors.

Changing Tastes or Preferences

From 1980 to 2012, the per-person consumption of chicken by Americans rose from 33 pounds per year to 81 pounds per year, and consumption of beef fell from 77 pounds per year to 57 pounds per year, according to the U.S. Department of Agriculture (USDA). Changes like these are largely due to shifts in taste, which change the quantity of a good demanded at every price: That is, they shift the demand curve for that good—rightward for chicken and leftward for beef.

Changes in the Composition of the Population

The proportion of elderly citizens in the United States population is rising. It rose from 9.8 percent in 1970 to 12.6 percent in 2000 and will be a projected (by the U.S. Census Bureau) 20 percent of the population by 2030. A society with relatively more children, like the United States in the 1960s, will have greater demand for goods and services like tricycles and day care facilities. A society with relatively more elderly persons, as the United States is projected to have by 2030, has a higher demand for nursing homes and hearing aids. Similarly, changes in the size of the population can affect the demand for housing and many other goods. Each of these changes in demand will be shown as a shift in the demand curve.

Changes in the Prices of Related Goods

The demand for a product can also be affected by changes in the prices of related goods such as substitutes or complements. A **substitute** is a good or service that can be used in place of another good or service. As electronic books, like this one, become more available, you would expect to see a decrease in demand for traditional printed books. A lower price for a substitute decreases demand for the other product. For example, in recent years as the price of tablet computers has fallen, the quantity demanded has increased (because of the law of demand). Since people are purchasing tablets, there has been a decrease in demand for laptops, which can be shown graphically as a leftward shift in the demand curve for laptops. A higher price for a substitute good has the reverse effect.

Other goods are **complements** for each other, meaning that the goods are often used together, because consumption of one good tends to enhance consumption of the other. Examples include breakfast cereal and milk; notebooks and pens or pencils, golf balls and golf clubs; gasoline and sport utility vehicles; and the five-way combination of bacon, lettuce, tomato, mayonnaise, and bread. If the price of golf clubs rises, since the quantity of golf clubs demanded falls (because of the law of demand), demand for a complement good like golf balls decreases, too. Similarly, a higher price for skis would shift the demand curve for a complement good like ski resort trips to the left, while a lower price for a complement has the reverse effect.

Changes in Expectations About Future Prices or Other Factors That Affect Demand

While it is clear that the price of a good affects the quantity demanded, it is also true that expectations about the future price (or expectations about tastes and preferences, income, and so on) can affect demand. For example, if people hear that a hurricane is coming, they may rush to the store to buy flashlight batteries and bottled water. If people learn that the price of a good like coffee is likely to rise in the future, they may head for the store to stock up on coffee now. These changes in demand are shown as shifts in the curve. Therefore, a **shift in demand** happens when a change in some economic factor (other than the current price) causes a different quantity to be demanded at every price.

54. Worked Example: Shift in Demand

Shift in Demand Due to Income Increase

A shift in demand means that at any price (and at every price), the quantity demanded will be different than it was before. Following is a graphic illustration of a shift in demand due to an income increase.

Step 1. Draw the graph of a demand curve for a normal good like pizza. Pick a price (like P_0). Identify the corresponding Q_0. An example is shown in Figure 1.

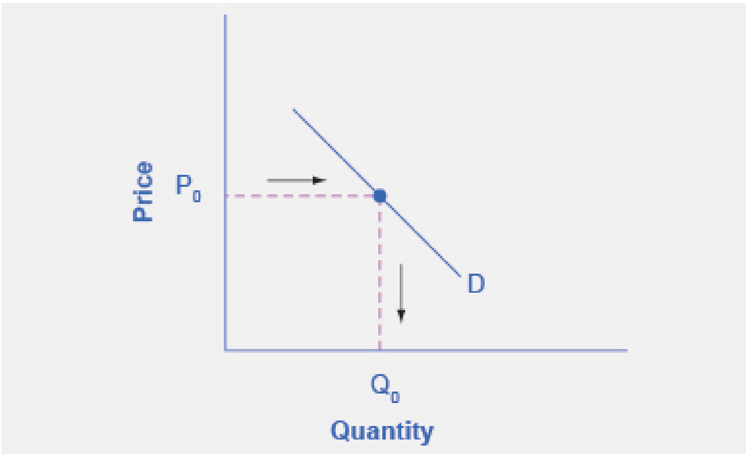


Figure 1. Demand Curve. *A demand curve can be used to identify how much consumers would buy at any given price.*

Step 2. Suppose income increases. As a result of the change, are consumers going to buy more or less pizza? The answer is more. Draw a dotted horizontal line from the chosen price, through the original quantity demanded, to the new point with the new Q_1. Draw a dotted vertical line down to the horizontal axis and label the new Q_1. An example is provided in Figure 2.

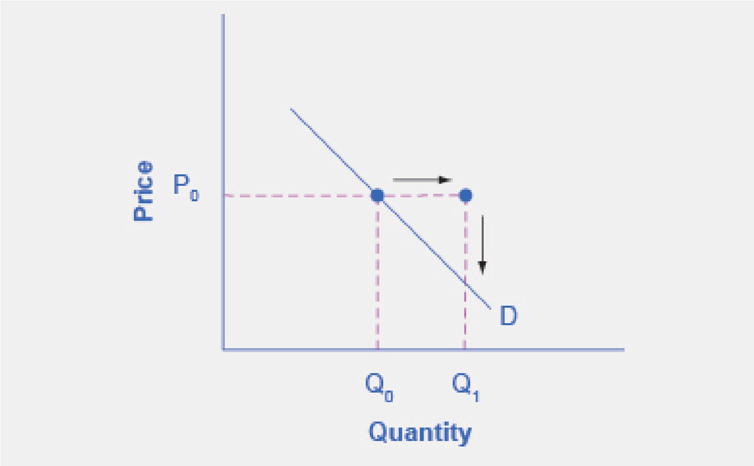


Figure 2. Demand Curve with Income Increase. *With an increase in income, consumers will purchase larger quantities, pushing demand to the right.*

Step 3. Now, shift the curve through the new point. You will see that an increase in income causes an upward (or rightward) shift in the demand curve, so that at any price, the quantities demanded will be higher, as shown in Figure 3.

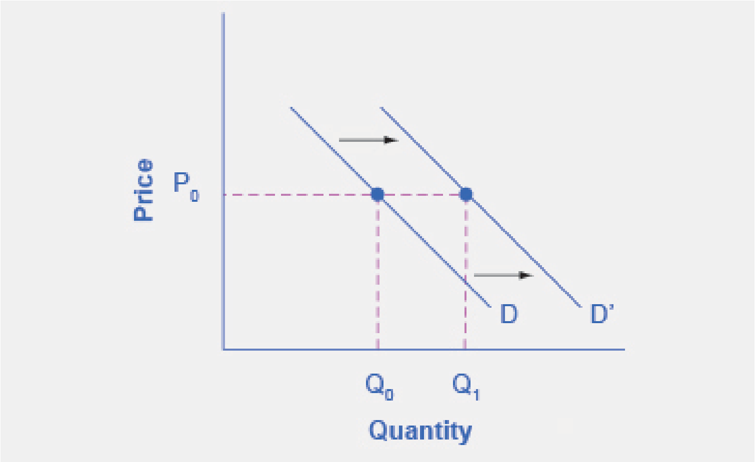
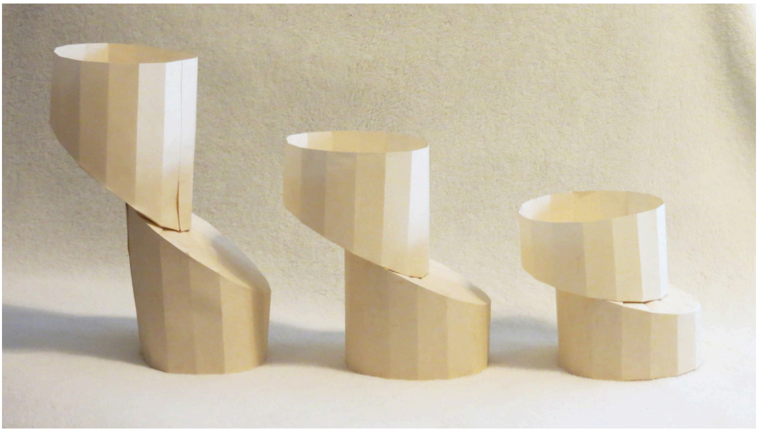


Figure 3. Demand Curve Shifted Right. *With an increase in income, consumers will purchase larger quantities, pushing demand to the right, and causing the demand curve to shift right.*

55. Reading: Summary of Factors That Change Demand



Six factors that can shift demand curves are summarized in Figure 1, below. The direction of the arrows indicates whether the demand curve shifts represent an increase in demand or a decrease in demand. Notice that a change in the price of the good or service itself is not listed among the factors that can shift a demand curve. A change in the price of a good or service causes a movement along a specific demand curve, and it typically leads to some change in the quantity demanded, but it does not shift the demand curve.

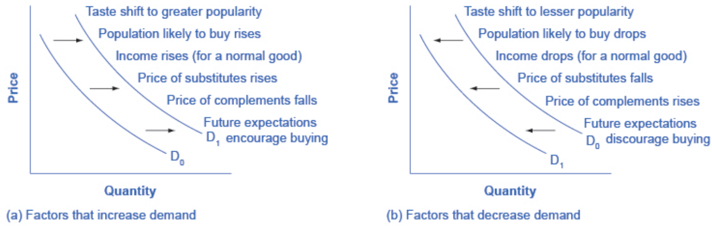


Figure 1. Factors That Shift Demand Curves *(a) A list of factors that can cause an increase in demand from D0 to D1. (b) The same factors, if their direction is reversed, can cause a decrease in demand from D0 to D1.*

56. Simulation: Demand for Food Trucks

Try It

Play the simulation below multiple times to see how different choices lead to different outcomes. All simulations allow unlimited attempts so that you can gain experience applying the concepts.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=77

57. Self Check: Demand

Check Your Understanding

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the four Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1559

58. Outcome: Supply

What you'll learn to do: explain the determinants of supply

So far you've learned about the role of demand in economics—which is the consumer side of the story. In this section, you'll learn about the producer side of economics to see what factors impact the amount of goods supplied in a market. For example, suppose the global price of petroleum falls significantly. What do you think will happen to the supply of gasoline? How are supply and price connected? In this section you'll examine the law of supply and see why this counterpart to "demand" is also essential to understanding economics.

The specific things you'll learn in this section include the following:

- Explain the law of supply
- Explain a supply curve
- Create a supply curve using a data set
- Describe the differences between changes in supply and changes in quantity supplied
- Explain the impact of factors that change supply

Learning Activities

The learning activities for this section include the following:

- Reading: What Is Supply?
- Reading: Factors Affecting Supply

- Worked Example: Shift in Supply
- Reading: Summary of Factors That Change Supply
- Simulation: Supply of Food Trucks
- Self Check: Supply

59. Reading: What Is Supply?

Supply of Goods and Services

When economists talk about **supply**, they mean the amount of some good or service a producer is willing to supply at each price. Price is what the producer receives for selling one unit of a good or service. A rise in price almost always leads to an increase in the quantity supplied of that good or service, while a fall in price will decrease the quantity supplied. When the price of gasoline rises, for example, it encourages profit-seeking firms to take several actions: expand exploration for oil reserves; drill for more oil; invest in more pipelines and oil tankers to bring the oil to plants where it can be refined into gasoline; build new oil refineries; purchase additional pipelines and trucks to ship the gasoline to gas stations; and open more gas stations or keep existing gas stations open longer hours. Economists call this positive relationship between price and quantity supplied—that a higher price leads to a higher quantity supplied and a lower price leads to a lower quantity supplied—the **law of supply**. The law of supply, like the law of demand, assumes that all other variables that affect supply (to be explained in the next reading) are held equal.

Supply vs. Quantity Supplied

In economic terminology, supply is not the same as quantity supplied. When economists refer to supply, they mean the relationship between a range of prices and the quantities supplied at those prices, a relationship that can be illustrated with a supply curve or a supply schedule. When economists refer to **quantity supplied**, they mean only a certain point on the supply curve, or one quantity on the supply schedule. In short, supply refers to the curve, and quantity supplied refers to the (specific) point on the curve.

Figure 1, below, illustrates the law of supply, again using the market for gasoline as an example. Like demand, supply can be illustrated using a table or a graph. A **supply schedule** is a table—like Table 1, below—that shows the quantity supplied at a range of different prices. Again, price is measured in dollars per gallon of gasoline, and quantity demanded is measured in millions of gallons. A **supply curve** is a graphic illustration of the relationship between price, shown on the vertical axis, and quantity, shown on the horizontal axis. You can see from this curve (Figure 1) that as the price rises, quantity supplied also increases and vice versa. The supply schedule and the supply curve are just two different ways of showing the same information. Notice that the horizontal and vertical axes on the graph for the supply curve are the same as for the demand curve.

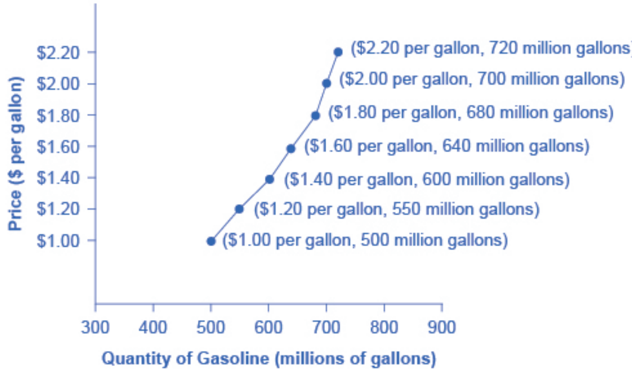


***Figure 1.* A Supply Curve for Gasoline**

Table 1. Price and Supply of Gasoline

Price (per gallon)	Quantity Supplied (millions of gallons)
$1.00	500
$1.20	550
$1.40	600
$1.60	640
$1.80	680
$2.00	700
$2.20	720

The shape of supply curves will vary somewhat according to the product: steeper, flatter, straighter, or curved. Nearly all supply curves, however, share a basic similarity: They slope up from left to right and illustrate the law of supply. As the price rises, say, from $1.00 per gallon to $2.20 per gallon, the quantity supplied increases from 500 gallons to 720 gallons. Conversely, as the price falls, the quantity supplied decreases.

The Law of Supply Video

The law of supply states that more of a good will be provided the higher its price; less will be provided the lower its price, *ceteris paribus*. There is a direct relationship between price and quantity supplied.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=80#oembed-1

60. Reading: Factors Affecting Supply

How Production Costs Affect Supply

A supply curve shows how quantity supplied will change as the price rises and falls, assuming *ceteris paribus*, so that no other economically relevant factors are changing. If other factors relevant to supply do change, then the entire supply curve will shift. Just as a shift in demand is represented by a change in the quantity demanded at every price, a **shift in supply** means a change in the quantity supplied at every price.

In thinking about the factors that affect supply, remember what motivates firms: profits, which are the difference between revenues and costs. Goods and services are produced using combinations of labor, materials, and machinery, or what we call **inputs** (also called **factors of production)**. If a firm faces lower costs of production, while the prices for the good or service the firm produces remain unchanged, a firm's profits go up. When a firm's profits increase, it's more motivated to produce **output** (goods or services), since the more it produces the more profit it will earn. So, when costs of production fall, a firm will tend to supply a larger quantity at any given price for its output. This can be shown by the supply curve shifting to the right.

Take, for example, a messenger company that delivers packages around a city. The company may find that buying gasoline is one of its main costs. If the price of gasoline falls, then the company will find it can deliver packages more cheaply than before. Since lower costs correspond to higher profits, the messenger company may now supply more of its services at any given price. For example, given the lower gasoline prices, the company can now serve a greater area, and increase its supply.

Conversely, if a firm faces higher costs of production, then it will earn lower profits at any given selling price for its products. As a result, a higher cost of production typically causes a firm to supply a smaller quantity at any given price. In this case, the supply curve shifts to the left.

Consider the supply for cars, shown by curve S_0 in Figure 1, below. Point J indicates that if the price is $20,000, the quantity supplied will be 18 million cars. If the price rises to $22,000 per car, *ceteris paribus*, the quantity supplied will rise to 20 million cars, as point K on the S_0 curve shows. The same information can be shown in table form, as in Table 1.

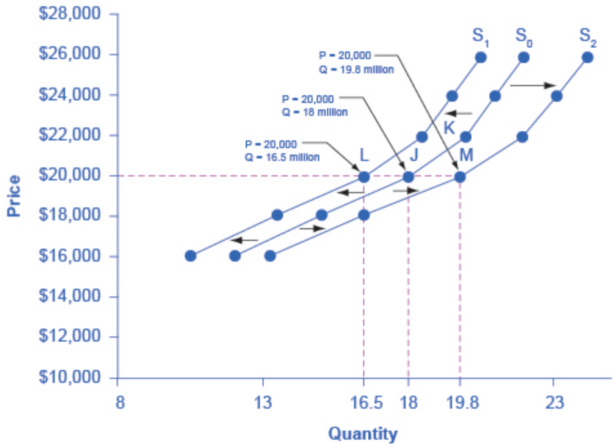


Figure 1. Shifts in Supply: A Car Example

Table 1. Price and Shifts in Supply: A Car Example

Price	Decrease to S_1	Original Quantity Supplied S_0	Increase to S_2
$16,000	10.5 million	12.0 million	13.2 million
$18,000	13.5 million	15.0 million	16.5 million
$20,000	16.5 million	18.0 million	19.8 million
$22,000	18.5 million	20.0 million	22.0 million
$24,000	19.5 million	21.0 million	23.1 million

Now imagine that the price of steel—an important component in vehicle manufacturing—rises, so that producing a car has become more expensive. At any given price for selling cars, car manufacturers will react by supplying a lower quantity. This can be shown graphically as a leftward shift of supply, from S_0 to S_1, which indicates that at any given price, the quantity supplied decreases. In this example, at a price of $20,000, the quantity supplied decreases from 18 million on the original supply curve (S_0) to 16.5 million on the supply curve S_1, which is labeled as point L.

Conversely, if the price of steel decreases, producing a car becomes less expensive. At any given price for selling cars, car manufacturers can now expect to earn higher profits, so they will supply a higher quantity. The shift of supply to the right, from S_0 to S_2, means that at all prices, the quantity supplied has increased. In this example, at a price of $20,000, the quantity supplied increases from 18 million on the original supply curve (S_0) to 19.8 million on the supply curve S_2, which is labeled M.

Other Factors That Affect Supply

In the example above, we saw that changes in the prices of inputs in the production process will affect the cost of production and thus the supply. Several other things affect the cost of production, too, such as changes in weather or other natural conditions, new technologies for production, and some government policies.

The cost of production for many agricultural products will be affected by changes in natural conditions. For example, the area of northern China that typically grows about 60 percent of the country's wheat output experienced its worst drought in at least fifty years in the second half of 2009. A drought decreases the supply of agricultural products, which means that at any given price, a lower quantity will be supplied; conversely, especially good weather would shift the supply curve to the right.

When a firm discovers a new technology that allows it to produce at a lower cost, the supply curve will shift to the right, as well. For instance, in the 1960s a major scientific effort nicknamed the Green Revolution focused on breeding improved seeds for basic crops like wheat and rice. By the early 1990s, more than two-thirds of the wheat and rice in low-income countries around the world was grown with these Green Revolution seeds—and the harvest was twice as high per acre. A technological improvement that reduces costs of production will shift supply to the right, so that a greater quantity will be produced at any given price.

Government policies can affect the cost of production and the supply curve through taxes, regulations, and subsidies. For example, the U.S. government imposes a tax on alcoholic beverages that collects about $8 billion per year from producers. Taxes are treated as costs by businesses. Higher costs decrease supply for the reasons discussed above. Other examples of policy that can affect cost are the wide array of government regulations that require firms to spend money to provide a cleaner environment or a safer workplace; complying with regulations increases costs.

A government subsidy, on the other hand, is the opposite of a tax. A **subsidy** occurs when the government pays a firm directly or reduces the firm's taxes if the firm carries out certain actions. From the firm's perspective, taxes or regulations are an additional cost of production that shifts supply to the left, leading the firm to produce a lower quantity at every given price. Government subsidies reduce the cost of production and increase supply at every given price, shifting supply to the right.

61. Worked Example: Shift in Supply

Shift in Supply Due to Production-Cost Increase

We know that a supply curve shows the minimum price a firm will accept to produce a given quantity of output. What happens to the supply curve when the cost of production goes up? Following is an example of a shift in supply due to an increase in production cost.

Step 1. Draw a graph of a supply curve for pizza. Pick a quantity (like Q_0). If you draw a vertical line up from Q_0 to the supply curve, you will see the price the firm chooses. An example is shown in Figure 1.

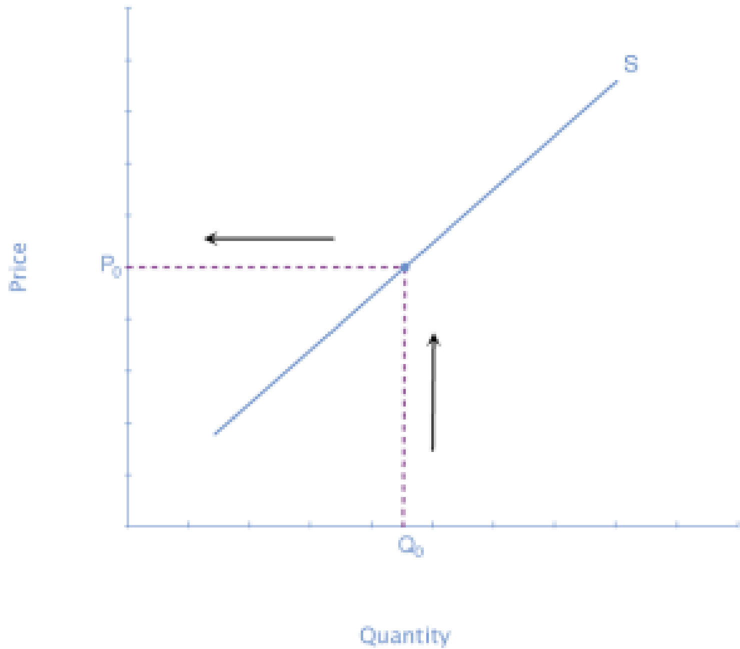


***Figure 1. Supply Curve**. The supply curve can be used to show the minimum price a firm will accept to produce a given quantity of output.*

Step 2. Why did the firm choose that price and not some other? One way to think about this is that the price is composed of two parts. The first part is the average cost of production: in this case, the cost of the pizza ingredients (dough, sauce, cheese, pepperoni, and so on), the cost of the pizza oven, the rent on the shop, and the wages of the workers. The second part is the firm's desired profit, which is determined, among other factors, by the profit margins in that particular business. If you add these two parts together, you get the price the firm wishes to charge. The quantity Q_0 and associated price P_0 give you one point on the firm's supply curve, as shown in Figure 2.

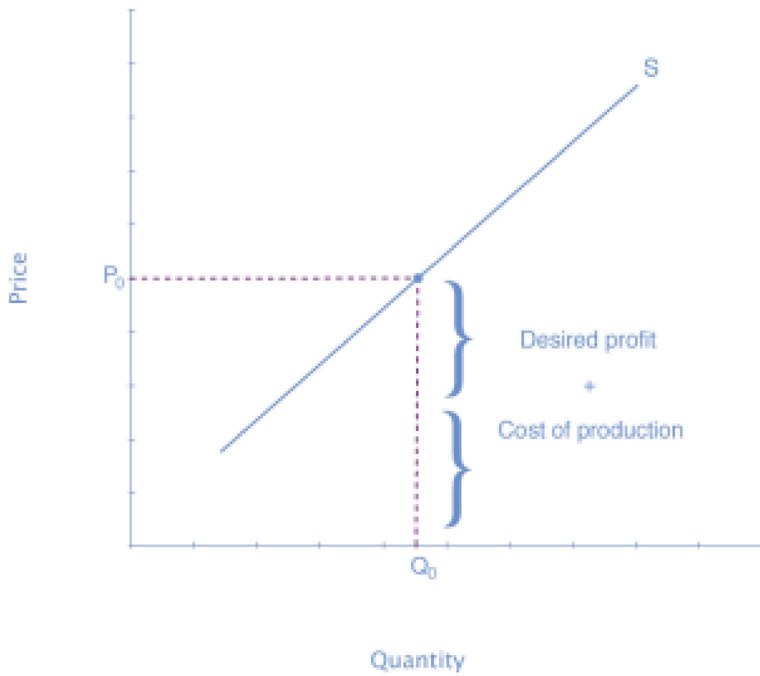


Figure 2. Setting Prices. *The cost of production and the desired profit equal the price a firm will set for a product.*

Step 3. Now, suppose that the cost of production goes up. Perhaps cheese has become more expensive by $0.75 per pizza. If that is true, the firm will want to raise its price by the amount of the increase in cost ($0.75). Draw this point on the supply curve directly above the initial point on the curve, but $0.75 higher, as shown in Figure 3.

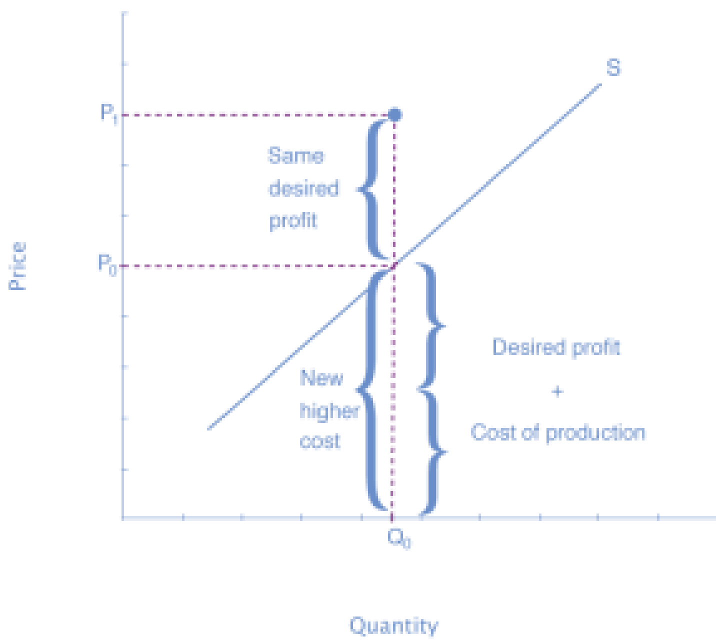


Figure 3. Increasing Costs Lead to Increasing Price. *Because the cost of production plus the desired profit equal the price a firm will set for a product, if the cost of production increases, the price for the product will also need to increase.*

Step 4. Shift the supply curve through this point. You will see that an increase in cost causes a leftward shift of the supply curve so that at any price, the quantities supplied will be smaller, as shown in Figure 4.

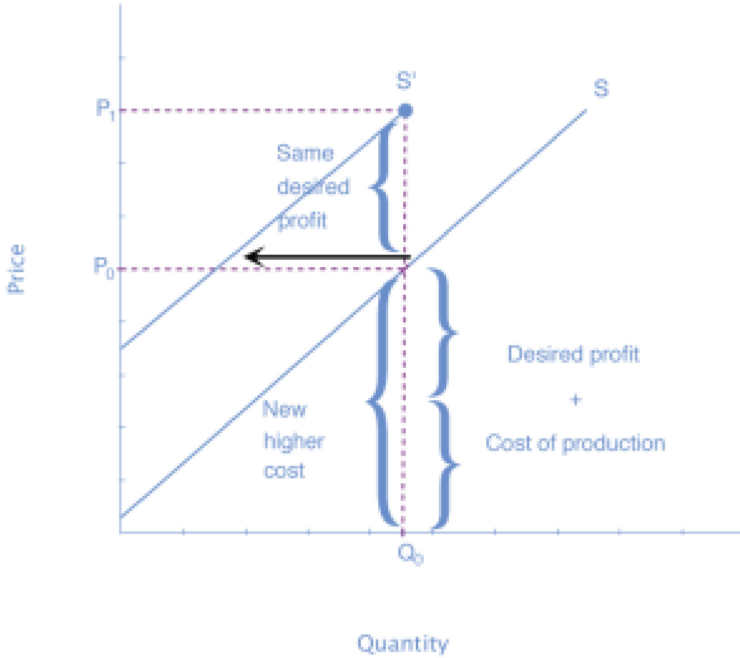


***Figure 4. Supply Curve Shifted Left**. When the cost of production increases, the supply curve shifts leftward to a new price level.*

62. Reading: Summary of Factors That Change Supply

Changes in the cost of inputs, natural disasters, new technologies, and the impact of government decisions all affect the cost of production. In turn, these factors affect how much firms are willing to supply at any given price.

Figure 1, below, summarizes factors that change the supply of goods and services. Notice that a change in the price of the product itself is not among the factors that shift the supply curve. Although a change in price of a good or service typically causes a change in quantity supplied or a movement along the supply curve for that specific good or service, it does not cause the supply curve itself to shift.

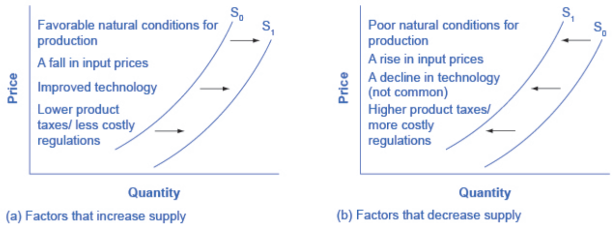


Figure 1. Factors That Shift Supply Curves. *(a) A list of factors that can cause an increase in supply from S0 to S1. (b) The same factors, if their direction is reversed, can cause a decrease in supply from S0 to S1.*

Because demand and supply curves appear on a two-dimensional diagram with only price and quantity on the axes, an unwary visitor to the land of economics might be fooled into believing that economics is about only four topics: demand, supply, price, and quantity. However, demand and supply are really "umbrella" concepts: demand covers all the factors that affect demand, and supply covers all the factors that affect supply. Factors other than price that affect demand and supply are included by using shifts in the demand or the supply curve. In this way, the two-dimensional demand and supply model becomes a powerful tool for analyzing a wide range of economic circumstances.

63. Simulation: Supply of Food Trucks

Try It

Play the simulation below multiple times to see how different choices lead to different outcomes. All simulations allow unlimited attempts so that you can gain experience applying the concepts.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=84

64. Self Check: Supply

Check Your Understanding

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the three Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1560

65. Outcome: Equilibrium

What you'll learn to do: explain and graphically illustrate market equilibrium, surplus, and shortage

In this section, you'll learn how supply and demand interact to determine the price in a market.

The specific things you'll learn in this section include the following:

- Define and explain equilibrium price and quantity
- Create a graph that illustrates equilibrium price and quantity
- Define and explain surpluses and shortages
- Create a graph that illustrates surpluses and shortages
- Describe how disequilibrium can create surpluses and shortages; explain how markets eliminate them

Learning Activities

The learning activities for this section include the following:

- Reading: Equilibrium, Surplus, and Shortage
- Video: Market Equilibrium
- Reading: Changes in Equilibrium
- Worked Example: Supply and Demand
- Simulation: Food Trucks and Changes in Equilibrium
- Self Check: Equilibrium

66. Reading: Equilibrium, Surplus, and Shortage

Demand and Supply

In order to understand market equilibrium, we need to start with the laws of demand and supply. Recall that the law of demand says that as price decreases, consumers demand a higher quantity. Similarly, the law of supply says that when price decreases, producers supply a lower quantity.

Because the graphs for demand and supply curves both have price on the vertical axis and quantity on the horizontal axis, the demand curve and supply curve for a particular good or service can appear on the same graph. Together, demand and supply determine the price and the quantity that will be bought and sold in a market. These relationships are shown as the demand and supply curves in Figure 1, which is based on the data in Table 1, below.

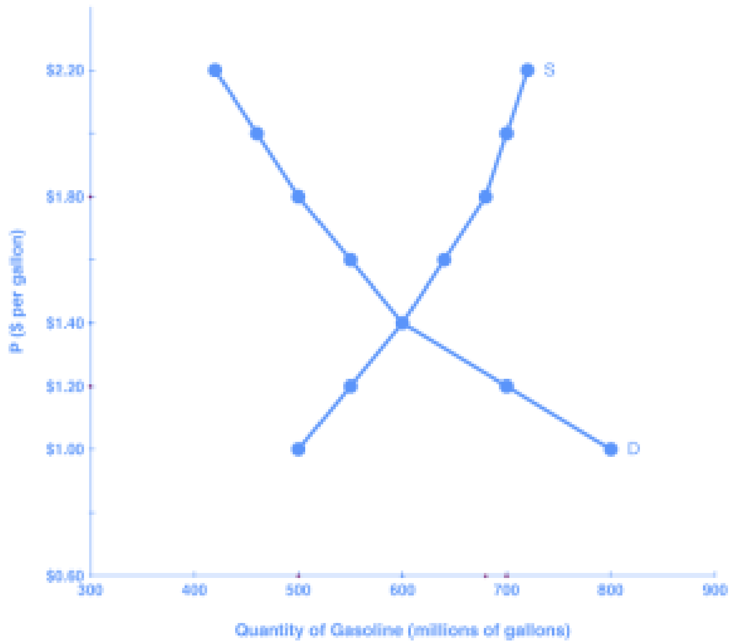


Figure 1. Demand and Supply for Gasoline

Table 1. Price, Quantity Demanded, and Quantity Supplied

Price (per gallon)	Quantity demanded (millions of gallons)	Quantity supplied (millions of gallons)
$1.00	800	500
$1.20	700	550
$1.40	**600**	**600**
$1.60	550	640
$1.80	500	680
$2.00	460	700
$2.20	420	720

If you look at either Figure 1 or Table 1, you'll see that, at most prices, the amount that consumers want to buy (which we call quantity demanded) is different from the amount that producers want to sell (which we call quantity supplied). What does it mean when the quantity demanded and the quantity supplied aren't the same? Answer: a surplus or a shortage.

Surplus or Excess Supply

Let's consider one scenario in which the amount that producers want to sell doesn't match the amount that consumers want to buy. Suppose that a market produces *more* than the quantity demanded. Let's use our example of the price of a gallon of gasoline. Imagine that the price of a gallon of gasoline were $1.80 per gallon. This price is illustrated by the dashed horizontal line at the price of $1.80 per gallon in Figure 2, below.

The graph shows the demand and supply curves for gasoline; the two curves intersect at the point of equilibrium. The lines resemble an "X." Price is shown on the y-axis, and quantity of gasoline is shown on the x-axis. The region above the equilibrium point (where the curves intersect) indicates excess supply, or surplus

Figure 2. Demand and Supply for Gasoline: Surplus

At this price, the quantity demanded is 500 gallons, and the quantity of gasoline supplied is 680 gallons. You can also find these numbers in Table 1, above. Now, compare quantity demanded and quantity supplied at this price. Quantity supplied (680) is greater than quantity demanded (500). Or, to put it in words, the amount that producers want to sell is greater than the amount that consumers want to buy. We call this a situation of **excess supply** (since Qs > Qd) or a **surplus**. Note that whenever we compare supply and demand, it's in the context of a specific price—in this case, $1.80 per gallon.

With a surplus, gasoline accumulates at gas stations, in tanker trucks, in pipelines, and at oil refineries. This accumulation puts pressure on gasoline sellers. If a surplus remains unsold, those firms involved in making and selling gasoline are not receiving enough cash to pay their workers and cover their expenses. In this situation, some producers and sellers will want to cut prices, because it is better to sell at a lower price than not to sell at all. Once some sellers start cutting prices, others will follow to avoid losing sales. These price reductions will, in turn, stimulate a higher quantity demanded.

How far will the price fall? Whenever there is a surplus, the price will drop until the surplus goes away. When the surplus is eliminated, the quantity supplied just equals the quantity demanded—that is, the amount that producers want to sell exactly equals the amount that consumers want to buy. We call this **equilibrium**, which means "balance." In this case, the equilibrium occurs at a price of $1.40 per gallon and at a quantity of 600 gallons. You can see this in Figure 2 (and Figure 1) where the supply and demand curves cross. You can also find it in Table 1 (the numbers in **bold**).

Equilibrium: Where Supply and Demand Intersect

When two lines on a diagram cross, this intersection usually means something. On a graph, the point where the supply curve (S) and the demand curve (D) intersect is the **equilibrium**. The **equilibrium price** is the only price where the desires of consumers and the desires of producers agree—that is, where the amount of the product that consumers want to buy (quantity demanded) is equal to the amount producers want to sell (quantity supplied). This mutually desired amount is called the **equilibrium quantity**. At any other price, the quantity demanded does not equal the quantity supplied, so the market is not in equilibrium at that price.

If you have only the demand and supply schedules, and no graph, you can find the equilibrium by looking for the price level on the tables where the quantity demanded and the quantity supplied are equal (again, the numbers in **bold** in Table 1 indicate this point).

Finding Equilibrium with Algebra

We've just explained two ways of finding a market equilibrium: by looking at a table showing the quantity demanded and supplied at different prices, and by looking at a graph of demand and supply. We can also identify the equilibrium with a little algebra if we have equations for the supply and demand curves.

Let's practice solving a few equations that you will see later in the course. Right now, we are only going to focus on the math. Later you'll learn why these models work the way they do, but let's start by focusing on solving the equations.

Suppose that the demand for soda is given by the following equation:

$$Qd = 16–2P$$

where Qd is the amount of soda that consumers want to buy (i.e., quantity demanded), and P is the price of soda.

Suppose the supply of soda is

$$Qs = 2 + 5P$$

where Qs is the amount of *t* that producers will supply (i.e., quantity supplied).

Finally, suppose that the soda market operates at a point where supply equals demand, or

$$Qd = Qs$$

We now have a system of three equations and three unknowns (Qd, Qs, and P), which we can solve with algebra. Since $Qd = Qs$, we can set the demand and supply equation equal to each other:

$$Qd = Qs$$
$$16 - 2P = 2 + 5P$$

Step 1: Isolate the variable by adding 2P to both sides of the equation, and subtracting 2 from both sides.

$$16 - 2P = 2 + 5P$$
$$-2 + 2P = -2 + 2P$$
$$14 = 7P$$

Step 2: Simplify the equation by dividing both sides by 7.

$$\frac{14}{7} = \frac{7P}{7}$$
$$2 = P$$

The price of each soda will be $2. Now we want to understand the amount of soda that consumers want to buy, or the quantity demanded, at a price of $2.

Remember, the formula for quantity demanded is the following:

$$Qd = 16 - 2P$$

Taking the price of $2, and plugging it into the demand equation, we get

$$Qd = 16–2(2)$$
$$Qd = 16–4$$
$$Qd = 12$$

So, if the price is $2 each, *consumers* will purchase 12. How much will producers supply, or what is the quantity supplied? Taking the price of $2, and plugging it into the equation for quantity supplied, we get the following:

$$Qs = 2 + 5P$$
$$Qs = 2 + 5(2)$$
$$Qs = 2 + 10$$
$$Qs = 12$$

Now, if the price is $2 each, *producers* will supply 12 sodas. This means that we did our math correctly, since $Qd = Qs$ and both Qd and Qs are equal to 12.

Shortage or Excess Demand

Let's return to our gasoline problem. Suppose that the price is $1.20 per gallon, as the dashed horizontal line at this price in Figure 3, below, shows. At this price, the quantity demanded is 700 gallons, and the quantity supplied is 550 gallons.

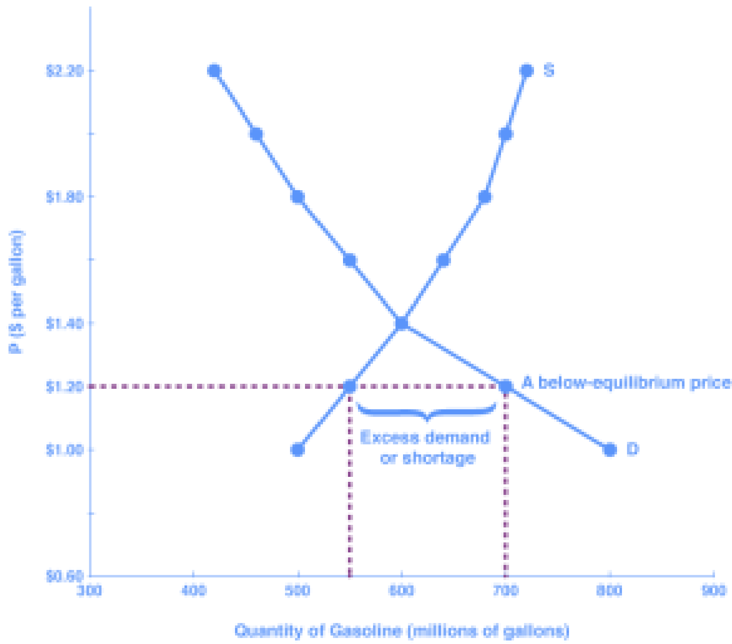


Figure 3. Demand and Supply for Gasoline: Shortage

Quantity supplied (550) is less than quantity demanded (700). Or, to put it in words, the amount that producers want to sell is less than the amount that consumers want to buy. We call this a situation of **excess demand** (since Qd > Qs) or a **shortage**.

In this situation, eager gasoline buyers mob the gas stations, only to find many stations running short of fuel. Oil companies and gas stations recognize that they have an opportunity to make higher profits by selling what gasoline they have at a higher price. These price increases will stimulate the quantity supplied and reduce the quantity demanded. As this occurs, the shortage will decrease. How far will the price rise? The price will rise until the shortage is eliminated and the quantity supplied equals quantity demanded. In other words, the market will be in equilibrium again. As before, the equilibrium occurs at a price of $1.40 per gallon and at a quantity of 600 gallons.

Generally any time the price for a good is *below* the equilibrium level, incentives built into the structure of demand and supply will create pressures for the price to rise. Similarly, any time the price for a good is *above* the equilibrium level, similar pressures will generally cause the price to fall.

As you can see, the quantity supplied or quantity demanded in a free market will correct over time to restore balance, or equilibrium.

Equilibrium and Economic Efficiency

Equilibrium is important to create both a balanced market and an efficient market. If a market is at its equilibrium price and quantity, then it has no reason to move away from that point, because it's balancing the quantity supplied and the quantity demanded. However, if a market is not at equilibrium, then economic pressures arise to move the market toward the equilibrium price and equilibrium quantity. This happens either because there is more supply than what the market is demanding or because there is more demand than the market is supplying. This balance is a natural function of a free-market economy.

Also, a competitive market that is operating at equilibrium is an efficient market. Economist typically define efficiency in this way: when it is impossible to improve the situation of one party without imposing a cost on another. Conversely, if a situation is inefficient, it becomes possible to benefit at least one party without imposing costs on others.

Efficiency in the demand and supply model has the same basic meaning: The economy is getting as much benefit as possible from its scarce resources, and all the possible gains from trade have been achieved. In other words, the optimal amount of each good and service is being produced and consumed.

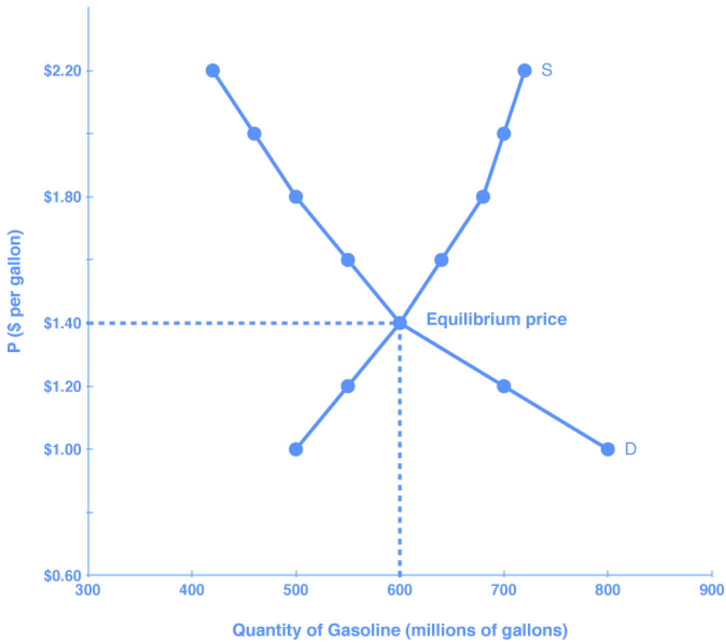


Figure 4. Demand and Supply for Gasoline: Equilibrium

67. Video: Market Equilibrium

https://youtu.be/W5nHpAn6FvQ?t=1s

Equilibrium occurs at the point where quantity supplied = quantity demanded.

68. Reading: Changes in Equilibrium

The Four-Step Process

Let's begin this discussion with a single economic event. It might be an event that affects demand, like a change in income, population, tastes, prices of substitutes or complements, or expectations about future prices. It might be an event that affects supply, like a change in natural conditions, input prices, technology, or government policies that affect production. How does an economic event like one of these affect equilibrium price and quantity? We'll investigate this question using a four-step process.

Step 1. Draw a demand and supply model before the economic change took place. Creating the model requires four standard pieces of information: the law of demand, which tells us the slope of the demand curve; the law of supply, which gives us the slope of the supply curve; the shift variables for demand; and the shift variables for supply. From this model, find the initial equilibrium values for price and quantity.

Step 2. Decide whether the economic change being analyzed affects demand or supply. In other words, does the event refer to something in the list of demand factors or supply factors?

Step 3. Decide whether the effect on demand or supply causes the curve to shift to the right or to the left, and sketch the new demand or supply curve on the diagram. In other words, does the event increase or decrease the amount consumers want to buy or producers want to sell?

Step 4. Identify the new equilibrium, and then compare the original equilibrium price and quantity to the new equilibrium price and quantity.

Newspapers and the Internet

According to the Pew Research Center for People and the Press, more and more people, especially younger people, are getting their news from online and digital sources. The majority of U.S. adults now own smartphones or tablets, and most of those Americans say they use them in part to get the news. From 2004 to 2012, the share of Americans who reported getting their news from digital sources increased from 24 percent to 39 percent. How has this trend affected consumption of print news media and radio and television news? Figure 1 and the text below illustrate the four-step analysis used to answer this question.

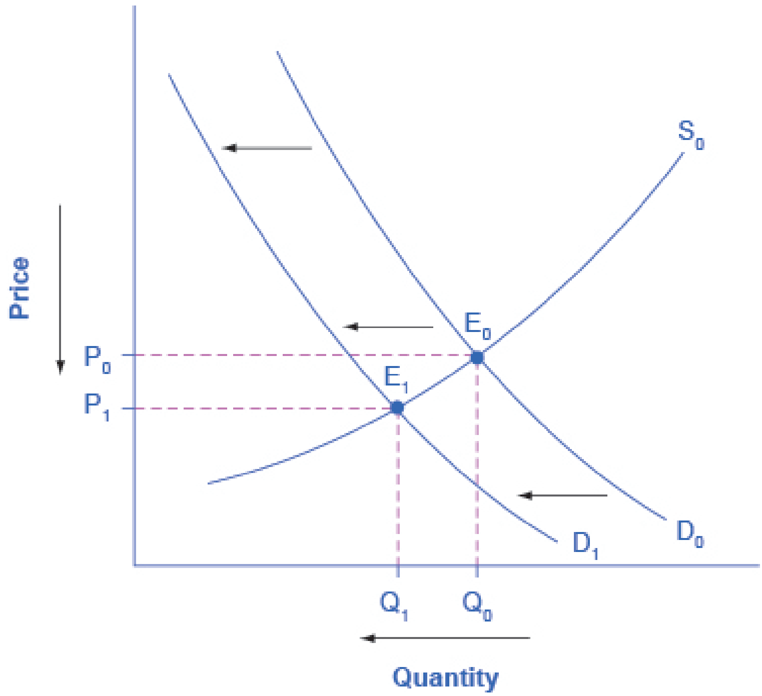


Figure 1. The Print News Market: A Four-Step Analysis

Step 1. Develop a demand and supply model to think about what the market looked like before the event. The demand curve D0 and the supply curve S0 show the original relationships. In this case, the analysis is performed without specific numbers on the price and quantity axis.

Step 2. Did the change described affect supply or demand? A change in tastes, from traditional news sources (print, radio, and television) to digital sources, caused a change in demand for the former.

Step 3. Was the effect on demand positive or negative? A shift to digital news sources will tend to mean a lower quantity demanded of traditional news sources at every given price, causing the demand curve for print and other traditional news sources to shift to the left, from D0 to D1.

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium (E1) occurs at a lower quantity and a lower price than the original equilibrium (E0).

The decline in print news reading predates 2004. Print newspaper circulation peaked in 1973 and has declined since then due to competition from television and radio news. In 1991, 55 percent of Americans indicated that they got their news from print sources, while only 29 percent did so in 2012. Radio news has followed a similar path in recent decades, with the share of Americans getting their news from radio declining from 54 percent in 1991 to 33 percent in 2012. Television news has held its own during the last fifteen years, with the market share staying in the mid to upper fifties. What does this suggest for the future, given that two-thirds of Americans under thirty years old say they don't get their news from television at all?

69. Worked Example: Supply and Demand

Supply and Demand

The example we just considered showed a shift to the left in the demand curve, as a change in consumer preferences reduced demand for newspapers. Often changes in an economy affect both the supply and the demand curves, making it more difficult to assess the impact on the equilibrium price. Let's review one such example.

First, consider the following questions:

1. Suppose postal workers are successful in obtaining a pay raise from the U.S. Postal Service. Will this affect the supply or the demand for first-class mail? Why? Which determinant of demand or supply is being affected? Show graphically with before and after curves on the same axes. How will this change the equilibrium price and quantity of first-class mail?
2. How do you imagine the invention of email and text messaging affected the market for first-class mail? Why? Which determinant of demand or supply is being affected? Show graphically with before and after curves on the same axes. How

will this change the equilibrium price and quantity of first-class mail?
3. Suppose that postal workers get a pay raise *and* email and text messaging become common. What will the combined impact be on the equilibrium price and quantity of first-class mail?

In order to complete a complex analysis like this it's helpful to tackle the parts separately and then combine them, while thinking about possible interactions between the two parts that might affect the overall outcome.

Part 1: A Pay Raise for Postal Workers

A pay raise for postal workers would represent an increase in the cost of production for the Postal Service. Production costs are a factor that influences supply; thus, the pay raise should decrease the supply of first-class mail, shifting the supply curve vertically by the amount of the pay raise. Intuitively, all else held constant, the Postal Service would like to charge a higher price that incorporates the higher cost of production. That is not to say the higher price will stick. From the graph (Figure 1), it should be clear that at that higher price, the quantity supplied is greater than the quantity demanded—thus there would be a surplus, indicating that the price the Postal Service desires is not an equilibrium price. Or to put it differently, at the original price (P1), the decrease in supply causes a shortage driving up the price to a new equilibrium level (P2). Note that the price doesn't rise by the full amount of the pay increase. In short, a leftward shift in the supply curve causes a movement up the demand curve, resulting in a lower equilibrium quantity (Q2) and a higher equilibrium price (P2).

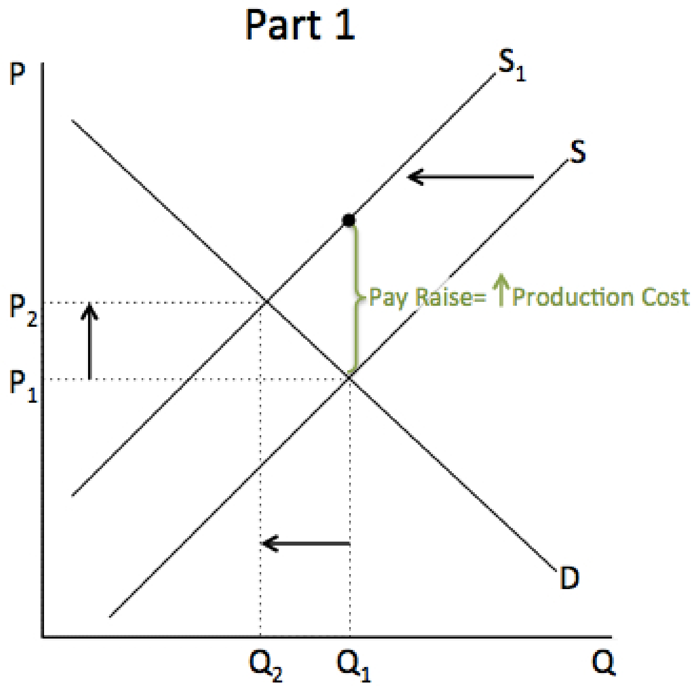


Figure 1

Part 2: The Effect of Email and Text Messaging

Since many people find email and texting more convenient than sending a letter, we can assume that tastes and preferences for first-class mail will decline. This decrease in demand is shown by a leftward shift in the demand curve and a movement along the supply curve, which creates a surplus in first-class mail at the original price (shown as P2). The shortage causes a decrease in the equilibrium price (to P3) and a decrease in the equilibrium quantity (to Q3). Intuitively, less demand for first-class mail leads to a lower equilibrium quantity and (ceteris paribus) a lower equilibrium price.

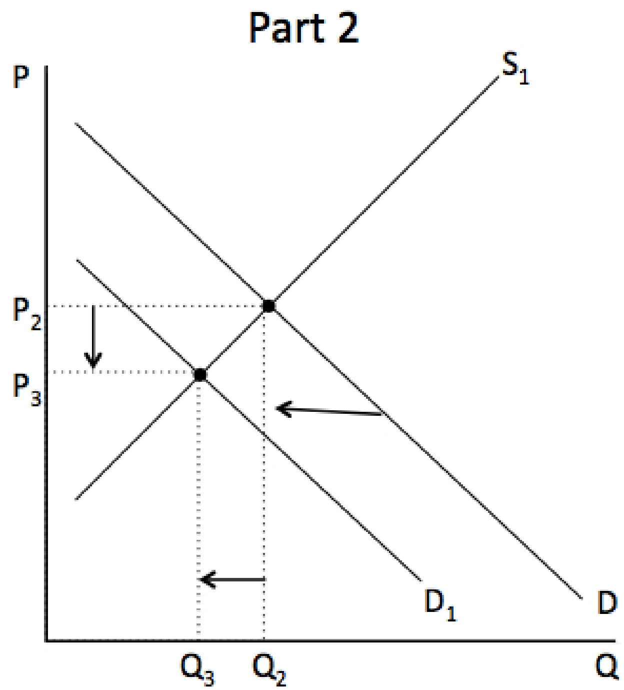


Figure 2

Part 3: Combining Factors

Parts 1 and 2 are straightforward, but when we put them together it becomes more complex. Think about it this way: In Part 1, the equilibrium quantity fell due to decreased supply. In Part 2, the equilibrium quantity also fell, this time due to the decreased demand. So putting the two parts together, we would expect to see the final equilibrium quantity (Q3) to be smaller than the original equilibrium quantity (Q1). So far, so good.

Now consider what happens to the price. In Part 1, the equilibrium price increased due to the reduction in supply. But in Part 2, the equilibrium price decreased due to the decrease in demand! What will happen to the equilibrium price? The net effect on price can't be determined without knowing which curve shifts more, demand or supply. The equilibrium price could increase, decrease, or stay the same. You just can't tell from graphical analysis alone.

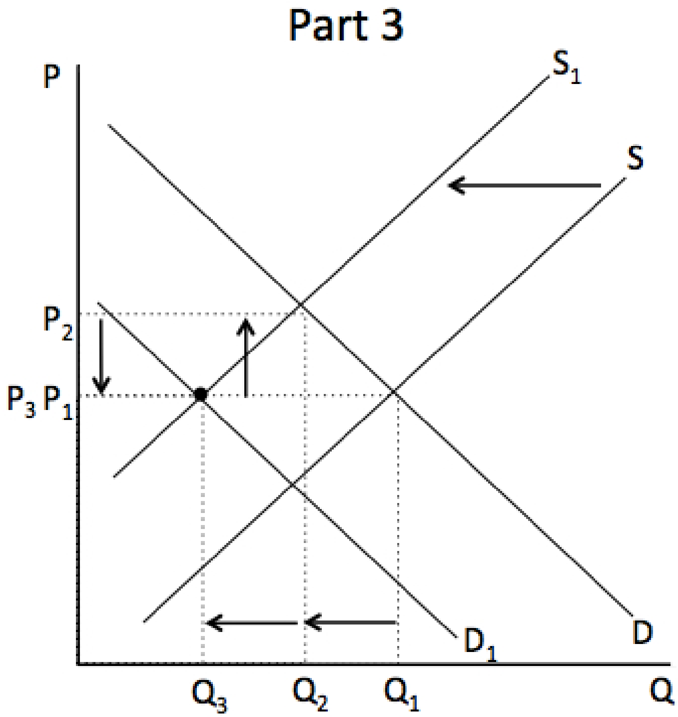


Figure 3

70. Simulation: Food Trucks and Changes in Equilibrium

Try It

Play the simulation below multiple times to see how different choices lead to different outcomes. All simulations allow unlimited attempts so that you can gain experience applying the concepts.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=91

71. Self Check: Equilibrium

Check Your Understanding

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the three Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1561

72. Problem Set: Supply and Demand 1

Test your understanding of the learning outcomes in this module by working through the following problems. These problems aren't graded, but they give you a chance to practice before taking the quiz.

If you'd like to try a problem again, you can click the link that reads, "Try another version of these questions."

Use the information provided in the first question for all of the questions in this problem set.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=93

73. Problem Set: Supply and Demand 2

Test your understanding of the learning outcomes in this module by working through the following problems. These problems aren't graded, but they give you a chance to practice before taking the quiz.

If you'd like to try a problem again, you can click the link that reads, "Try another version of these questions."

Use the information provided in the first question for all of the questions in this problem set.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=94

74. Problem Set: Supply and Demand 3

Test your understanding of the learning outcomes in this module by working through the following problems. These problems aren't graded, but they give you a chance to practice before taking the quiz.

If you'd like to try a problem again, you can click the link that reads, "Try another version of these questions."

Use the information provided in the first question for all of the questions in this problem set.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=95

75. Putting It Together: Supply and Demand

Summary

The goal of this module was to explain how demand and supply for a good or service determine prices and quantities bought and sold. In the process, you learned what factors influence demand and what factors influence supply. You learned how to:

- Describe and differentiate between the major economic systems
- Explain the determinants of demand
- Explain the determinants of supply
- Define and graphically illustrate market equilibrium, surplus, and shortage

Synthesis

You'll remember that we started this module by considering changes in global coffee prices. Let's focus specifically on the drought of 2014.

Figure 1. Coffee Prices. *Source: Trading Economics*

Now that we understand more about supply and demand, we can answer a few important questions: How does a drought impact supply? What impact will the quantity supplied have on the equilibrium price?

In 2014, the coffee regions of Brazil experienced a serious drought. The lack of rain in Brazil's coffee-growing region delayed the tree-flowering period, which spans October and November. When the trees don't flower, they don't produce coffee. Weather conditions also affect the pollination of coffee trees that have already flowered: Drought makes the blooms very delicate, which can cause them to fall off the tree. In 2014, the combined impact of these consequences meant a 13 percent drop in production from the previous year, to only 48 million 60-kilogram bags.[1]

1. http://www.wallstreetdaily.com/2014/10/21/coffee-prices-brazil-drought/

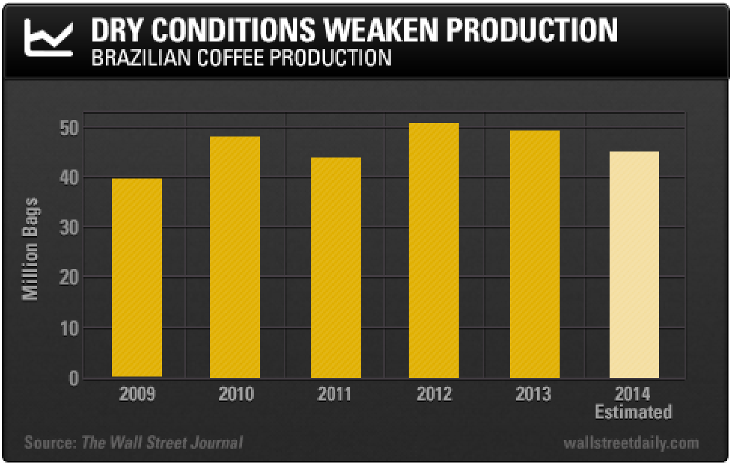


Figure 2

These are poor natural conditions for coffee growers, and they cause a reduction in the supply. Graphically, such a reduction means a shift to the left in the supply curve (shown in Figure 3, below), indicating that suppliers are providing less coffee at every price.

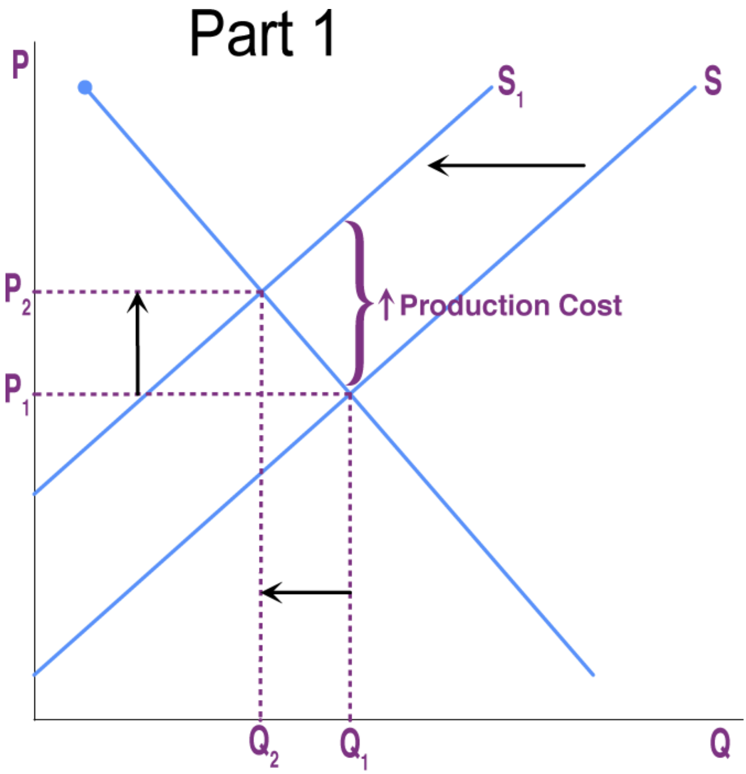


Figure 3

We can see that this shift in the supply curve will change the quantity supplied and the equilibrium price. At the original price (P1), the decrease in supply causes a shortage—more people want coffee at that low price than the suppliers are able to provide. This drives up the price to a new equilibrium level (P2). In short, a leftward shift in the supply curve causes a movement up the demand curve, resulting in a lower equilibrium quantity (Q2) and a higher equilibrium price (P2).

This impact is clear in an economic model like the graph above, but does it really affect consumers? Absolutely!—during this period, Starbucks raised its prices by 8 percent, and Folgers raised its prices by 9 percent. Coffee retailers were able to limit some of the impact of the rising coffee prices by drawing down their stock of green beans that were purchased before the drought and passing on some of the cost on to their customers as a higher price.[2]

It's very common to see the impact of drought and other natural factors on supply, equilibrium quantity, and equilibrium price. The following video provides a brief example in the United States.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=96#oembed-1

Will raising the price of coffee cause consumers to drink less coffee? How much of an impact does a price change have on demand? Now that you understand the basic principles of supply and demand, we're ready to take on these questions in the next module.

2. http://www.barrons.com/articles/rise-in-coffee-prices-nearing-peak-1444457073

76. Glossary: Supply and Demand

ceteris paribus other things being equal

command economy an economy where economic decisions are passed down from government authority and where resources are owned by the government; also called a "planned economy"

competitive market a large market in which there is a large number of buyers and sellers, so that no one can control the market price

free market a market in which the government does not intervene in any way

complements goods that are often used together, so consumption of one good tends to enhance consumption of the other

demand the amount of a good or service that consumers are willing and able to purchase at each price

demand curve a graphic representation of the relationship between price and quantity of a certain good or service demanded, with quantity on the horizontal axis and the price on the vertical axis

demand schedule a table that shows a range of prices for a certain good or service and the quantity demanded at each price

equilibrium when the quantity demanded is equal to the quantity supplied

equilibrium price the price where quantity demanded is equal to quantity supplied

equilibrium quantity the quantity at which quantity demanded and quantity supplied are equal for a certain price level

inferior good a product whose demand falls when income rises, and vice versa

inputs the combination of labor, materials, and machinery used to produce goods and services; *also called "factors of production"*

law of demand states that more of a good will be demanded (bought) the lower its price, and less of a good will be demanded (bought) the higher its price, ceteris paribus (other things being equal)

law of supply states that more of a good will be provided the higher its price; less will be provided the lower its price, *ceteris paribus* (other things being equal)

market interaction between potential buyers and sellers; a combination of demand and supply

market economy an economy in which economic decisions are decentralized, resources are owned by private individuals, and businesses supply goods and services based on demand

normal good a product whose demand rises when income rises, and vice versa

outputs goods or services

price what a buyer pays for a unit of a specific good or service

quantity demanded the total number of units of a good or service that consumers are willing to purchase at a given price

quantity supplied the total number of units of a good or service that producers are willing to sell at a given price

shift in demand when a change in some economic factor (other than price) causes a different quantity to be demanded at every price

shift in supply when a change in some economic factor (other than price) causes a different quantity to be supplied at every price

shortage at the existing price, the quantity demanded exceeds the quantity supplied; *also called "excess demand"*

subsidy when the government pays a firm directly or reduces the firm's taxes if the firm carries out certain actions

substitute a good that can replace another to some extent, so greater consumption of one good tends to mean less of the other

supply the amount of a good or service that a producer is willing to supply at each price

supply curve a line that shows the relationship between price and quantity supplied on a graph, with quantity supplied on the horizontal axis and price on the vertical axis

supply schedule a table that shows a range of prices for a good or service and the quantity supplied at each price

surplus at the existing price, the quantity supplied exceeds the quantity demanded; *also called "excess supply"*

77. Discussion: Supply and Demand

A key skill in economics is the ability to use the theory of supply and demand to analyze specific markets. In this week's discussion, you get a chance to demonstrate your ability to analyze the effects of several "shocks" to the market for coffee. Choose one of the three scenarios below.

Scenario 1: Suppose that, as part of an international trade agreement, the U.S. government reduces the tariff on imported coffee. Will this affect the supply or the demand for coffee? Why? Which determinant of demand or supply is being affected? Show graphically with before- and after-curves on the same axes. How will this change the equilibrium price and quantity of coffee? Explain your reasoning.

Scenario 2: Suppose the National Institutes of Health publishes a study finding that coffee drinking reduces the probability of getting colon cancer. How do you imagine this will affect the market for coffee? Why? Which determinant of demand or supply is being affected? Show graphically with before- and after-curves on the same axes. How will this change the equilibrium price and quantity of coffee? Explain your reasoning.

Scenario 3: Combine parts 1 and 2. Suppose that the U.S. government reduces the tariff on imported coffee, and a reputable study is published indicating that coffee drinkers have lower rates of colon cancer. What will the combined impact be on the equilibrium price and quantity of coffee? Explain your reasoning and show graphically. Make sure you think this through carefully!

PART V

MODULE: ELASTICITY

78. Why It Matters: Elasticity

Why measure how changes in price, income, or other factors affect the behavior of buyers and sellers?

Imagine going to your favorite coffee shop and having the waiter inform you that the pricing has changed. Instead of $3 for a cup of coffee, you will now be charged $2 for coffee, $1 for creamer, and $1 for your choice of sweetener. If you pay your usual $3 for a cup of coffee, you must choose between creamer and sweetener. If you want both, you now face an extra charge of $1. Sound absurd? Well, that's the situation Netflix customers found themselves in—facing a 60 percent price hike to retain the same service.

***Netflix On-Demand Media**. Netflix, Inc. is an American provider of on-demand Internet streaming media to many countries around the world, including the United States, and of flat rate DVD-by-mail in the United States.*

In early 2011, Netflix consumers paid about $10 a month for a package consisting of streaming video and DVD rentals. In July 2011, the company announced a packaging change. Customers wishing to retain both streaming video and DVD rental would be charged $15.98 per month, a price increase of about 60 percent. How would customers of the fourteen-year-old firm react? Would they abandon Netflix? Would the ease of access to other venues make a difference in how consumers responded to the Netflix price change? In this module, the answers to these questions—about the change in quantity with respect to a change in price—will be explored through a concept economists call *elasticity*.

Elasticity measures the behavioral response of economic agents in a given situation. Here are some examples:

- If a business raises its prices, will that have a large or small impact on demand?
- If you get a pay raise, how much more will you spend on food, clothing or entertainment?
- If hot dogs go on sale at the grocery store, how much additional mustard will consumers purchase?
- If the local Italian restaurant puts their pizza on sale, will the additional number of pizzas sold offset the discount on each item? In other words, will their sales revenues for pizza go up or down?

These are important real-world questions that we'll study in this module.

Also, before we get into the details: It can be easy to get hung up on the math of elasticity calculations. Learning to do these calculations is an important part of applying the elasticity principle, but the math will seem more intuitive if you master concept first: Understanding what elasticity means in a particular context will help you see what you're trying to calculate.

Learning Outcomes

- Explain the concept of elasticity
- Explain the price elasticity of demand and price elasticity of supply, and compute both using the midpoint method
- Explain and calculate other elasticities using common economic variables
- Explain the relationship between a firm's price elasticity of demand and total revenue

79. Outcome: Explaining Elasticity

What you'll learn to do: explain the concept of elasticity

Elasticity is an economics concept that measures the responsiveness of one variable to changes in another variable. For example, if you raise the price of your product, how will that affect your sales numbers? The variables in this question are *price* and *sales numbers*. Elasticity explains how much one variable, say sales numbers, will change in response to another variable, like the price of the product.

Mastering this concept resembles learning to ride a bike: It's tough at first, but when you get it, you won't forget. A rookie mistake is learning the calculations of elasticity but failing to grasp the idea. Make sure you don't do this! First take time to understand the concepts—then the calculations can be used simply to explain them in a numerical way.

The specific things you'll learn in this section include the following:

- Define elasticity
- Provide a practical example of elasticity
- Explain common factors that affect elasticity

Learning Activities

The learning activities for this section include the following:

- Reading: Introduction to Elasticity
- Video: Price Elasticity of Demand
- Reading: Examples of Elastic and Inelastic Demand

80. Reading: Introduction to Elasticity

Mr. Fantastic is elastic.

Think about the word *elastic*. It suggests that an item can be stretched. In economics, when we talk about **elasticity**, we're referring to how much something will stretch or change in response to another variable. Consider a rubber band, a leather strap, and a steel ring. If you pull on two sides of a rubber band (or Mr. Fantastic), the force will cause it to stretch a lot. If you use the same amount of force to pull on the ends of a leather strap, it will stretch somewhat, but not as much as the rubber band. If you pull on either side of a steel ring, applying the same amount of force, it probably won't stretch at all (unless you're very strong). Each of these materials (the rubber band, the leather strap, and the steel ring) displays a different amount of elasticity in response to being pulled, and all three fall somewhere on a continuum from very stretchy (elastic) to barely stretchy (inelastic).

There are different kinds of economic elasticity—for example, price elasticity of demand, price elasticity of supply, income elasticity of demand, and cross-price elasticity of demand—but the underlying property is always the same: how responsive or sensitive one thing is to a change in another thing.

Elastic and Inelastic Demand

Let's think about elasticity in the context of price and quantity demanded. We know from the law of demand that a rise in price will lead to a decrease in the quantity demanded. How much of a decrease? If a small change in price creates a large change in demand, then we would say that the demand is very elastic—that is, the demand is very sensitive to a change in price. If, on the other hand, a large change in price results in a very small change in demand, then we would say the demand is inelastic. Here's a way to keep this straight: Demand is **in**elastic when consumers are **in**sensitive to changes in price.

Consider the example of cigarette taxes and smoking rates—a classic example of inelastic demand. Cigarettes are taxed at both the state and federal level. As you might expect, the greater the amount of the tax increase, the fewer cigarettes are bought and consumed. Certain groups of cigarette smokers, such as teenage, minority, low-income, and casual smokers, are somewhat sensitive to changes in price: For every 10 percent increase in the price of a pack of cigarettes, the smoking rates drop about 7 percent. Notice that the demand doesn't decrease as much as the price increase, though. We can say, then, that the demand for cigarettes—at least among these groups—is relatively inelastic. Addicted adult smokers are even less sensitive to changes in price—most are willing to pay what it takes to support their smoking habit. We can say that their demand is even more inelastic.

You might think that elasticity isn't an important consideration when it comes to the price of cigarettes. Surely *any* reduction in the demand for cigarettes would be a good thing, right? Does it really matter whether whether the demand is elastic or inelastic? It does. The reason is that taxes on cigarettes serve two purposes: to raise tax revenue for government and to discourage smoking. On one hand, if a higher cigarette tax discourages consumption by quite a lot—meaning a very large reduction in cigarette sales—then the cigarette tax on each pack will not raise much revenue for the government. On the other hand, a higher cigarette tax that does not discourage consumption by much will actually raise more tax revenue for the government (but not have much impact on smoking rates). Thus, when Congress tries to calculate the effects of altering its cigarette tax, it must analyze *how much* the tax affects the quantity of cigarettes consumed. In other words, understanding the elasticity of cigarette demand is key to measuring the impact of taxes on government revenue AND public health.

This issue reaches beyond governments and taxes; every firm faces a similar challenge. Every time a firm considers raising the price that it charges, it needs to know how much a price increase will reduce the quantity of its product that is demanded. Conversely, when a firm puts its products on sale, it wants assurance that the lower price will lead to a significantly higher quantity demanded.

Mr. Fox is *inelastic.*

81. Video: Price Elasticity of Demand

This video provides a nice overview of the concept of elasticity and how it can be used. You'll learn how to calculate elasticities later in this module.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=103#oembed-1

82. Reading: Examples of Elastic and Inelastic Demand

Now that you have a general idea of what elasticity is, let's consider some of the factors that can help us predict whether demand for a product is likely to be elastic or inelastic. The following are important considerations:

- **Substitutes**: Price elasticity of demand is fundamentally about substitutes. If it's easy to find a substitute product when the price of a product increases, the demand will be more elastic. If there are few or no alternatives, demand will be less elastic.
- **Necessities vs. luxuries**: A necessity is something you absolutely must have, almost regardless of the price. A luxury

is something that would be nice to have, but it's not absolutely necessary. Consider the elasticity of demand for cookies. A buyer may enjoy a cookie, but it doesn't fulfill a critical need the way a snow shovel after a blizzard or a life-saving drug does. In general, the greater the necessity of the product, the less elastic, or more inelastic, the demand will be, because substitutes are limited. The more luxurious the product is, the more elastic demand will be.

- **Share of the consumer's budget**: If a product takes up a large share of a consumer's budget, even a small percentage increase in price may make it prohibitively expensive to many buyers. Take rental housing that's located close to downtown. Such housing might cost half of one's budget. A small percentage increase in rent could cause renters to relocate to cheaper housing in the suburbs, rather than reduce their spending on food, utilities, and other necessities. Therefore the larger the share of an item in one's budget, the more price elastic demand is likely to be. By contrast, suppose the local grocery store increased the price of toothpicks by 50 percent. Since toothpicks represent such a small part of a consumer's budget, even a significant increase in price is likely to have only a small effect on demand. Thus, the smaller the share of an item in one's budget, the more price inelastic demand is likely to be.
- **Short run versus long run:** Price elasticity of demand is usually lower in the short run, before consumers have much time to react, than in the long run, when they have greater opportunity to find substitute goods. Thus, demand is more price elastic in the long run than in the short run.
- **Competitive dynamics**: Goods that can only be produced by one supplier generally have inelastic demand, while products that exist in a competitive marketplace have elastic demand. This is because a competitive marketplace offers more options for the buyer.

With these considerations in mind, take a moment to see if you can figure out which of the following products have elastic demand and which have inelastic demand. It may be helpful to remember that when the buyer is **in**sensitive to price, demand is **in**elastic.

- Gasoline
- College textbooks
- Coffee
- Airline tickets
- Concert tickets
- Soft drinks
- Medical procedures

Inelastic Demand	Elastic Demand
Gasoline The demand for gasoline generally is fairly inelastic, especially in the short run. Car travel requires gasoline. The substitutes for car travel offer less convenience and control. Much car travel is necessary for people to move between activities and can't be reduced to save money. In the long run, though, more options are available, such as purchasing a more fuel-efficient car or choosing a job that is closer to where you work.	Gas from a Particular Station The demand for gasoline from any single gas station, or chain of gas stations, is highly elastic. Buyers can choose between comparable products based on price. There are often many stations in a small geographic area that are equally convenient.
Traditional Textbooks Generally an instructor assigns a textbook to the student, and the student who wants access to the learning materials must buy it, regardless of the price level. Because the student can't easily identify another textbook or resource that will ensure the same content and grade for the class, he has no substitutes and must buy the book at any price (or opt not to buy it at all).	New Textbook Distribution Channels Increasingly, students have new options to buy the same textbooks from different distribution channels at different price points. The introduction of new distribution channels is increasing options for buyers and having an impact on the price elasticity for publishers.
Specialty Coffee Drinks Many coffee shops have developed branded drinks and specialized experiences in order to reduce substitutes and build customer loyalty. While black coffee is available almost universally, there are few substitutes for a Starbucks Java Chip Frappuccino. Demand for such products is more inelastic.	Black Coffee Coffee is generally widely available at a level of quality that meets the needs of most buyers. The combination of a low price, relative to the buyer's spending power, and the fact that the product is sold by many different suppliers in a competitive market, make the demand highly elastic.

Inelastic Demand	Elastic Demand
Concert Tickets Only Taylor Swift can offer a Taylor Swift concert. She holds a monopoly on the creation and delivery of that experience. There is no substitute, and loyal fans are willing to pay for the experience. Because it is a scarce resource and the delivery is tightly controlled by a single provider, access to concerts has inelastic demand.	Airline Tickets Airline tickets are sold in a fiercely competitive market. Buyers can easily compare prices, and buyers experience the services provided by competitors as being very similar. Buyers can often choose not to travel if the cost is too high or substitute travel by car or train.
Medical Procedures Essential medical procedures have inelastic demand. The patient will pay what she can or what she must. In general, products that significantly affect health and well-being have inelastic demand.	Soft Drinks Soft drinks and many other nonessential items have highly elastic demand. There is competition among every brand and type of soda, and there are many substitutes for the entire category of soft drinks.

Self Check: Explaining Elasticity

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1564

83. Outcome: Calculating Price Elasticity

What you'll learn to do: explain the price elasticity of demand and price elasticity of supply, and compute both using the midpoint method

Remember, elasticity measures the responsiveness of one variable to changes in another variable. In the last section we looked at price elasticity of demand, or how much a change in price affects the quantity demanded. In this section we will look at both elasticity of demand and elasticity of supply. Supply can also be elastic, since a change in price will influence the quantity supplied.

In this section you'll also learn how to calculate elasticity, using the midpoint (or arc) method. This is the method favored by economists, since it gives very accurate results.

The specific things you'll learn in this section include the following:

- Define price elasticity of demand
- Define price elasticity of supply
- Calculate price elasticity using the midpoint method
- Mathematically differentiate between elastic, inelastic, and unitary elasticities of demand and supply
- Explain perfect elasticity and perfect inelasticity

Learning Activities

The learning activities for this section include the following:

- Reading: Calculating Percentage Changes and Growth Rates
- Reading: Calculating Price Elasticities
- Reading: Three Categories of Elasticity
- Reading: Polar Cases of Elasticity
- Self Check: Calculating Price Elasticity

84. Reading: Calculating Percentage Changes and Growth Rates

In order to measure elasticity, we need to calculate percentage change, also known as a **growth rate**. The formula for computing a growth rate is straightforward:

$$\text{Percentage change} = \frac{\text{Change in quantity}}{\text{Quantity}}$$

Suppose that a job pays $10 per hour. At some point, the individual doing the job is given a $2-per-hour raise. The percentage change (or growth rate) in pay is $\frac{\$2}{\$10} = 0.20 \text{ or } 20\%$.

Now, recall that we defined elasticity as the percentage change in something divided by the percentage change in something else.

Let's take the price elasticity of demand as an example. The price elasticity of demand is defined as the percentage change in quantity demanded divided by the percentage change in price:

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

There are two general methods for calculating elasticities: the point elasticity approach and the midpoint (or arc) elasticity approach. Elasticity looks at the percentage change in quantity demanded divided by the percentage change in price, but which quantity and which price should be the denominator in the percentage calculation? The point approach uses the initial price and initial quantity to measure percent change. This makes the math easier, but the more accurate approach is the midpoint approach, which uses the average price and average quantity over the price and quantity change. (These are the price and quantity halfway between the initial point and the final point. Let's compare the two approaches.

Suppose the quantity demanded of a product was 100 at one point on the demand curve, and then it moved to 103 at another point. The growth rate, or percentage change in quantity demanded, would be the change in quantity demanded $(103-100)$ divided by the average of the two quantities demanded $\frac{(103+100)}{2}$.

In other words, the growth rate:

$$= \frac{103-100}{(103+100)/2} = \frac{3}{101.5} = 0.0296 = 2.96\%\text{growth}$$

Note that if we used the point approach, the calculation would be:

$$\frac{(103\text{–}100)}{100} = 3\% \text{ growth}$$

This produces nearly the same result as the slightly more complicated midpoint method (3% vs. 2.96%). If you need a rough approximation, use the point method. If you need accuracy, use the midpoint method.

In this module you will often be asked to calculate the *percentage change* in the quantity, but keep in mind that this is the growth rate. That way as you work through the course and find a need to calculate the growth rates beyond elasticity, you will be well prepared.

85. Reading: Calculating Price Elasticities

Introduction

Remember, all elasticities measure the responsiveness of one variable to changes in another variable. In this section, we will focus on the price elasticity of demand and the price elasticity of supply, but the calculations for other elasticities are analogous.

Let's start with the definition:

> **Price elasticity of demand** is the percentage change in the quantity of a good or service demanded divided by the percentage change in the price.

The Midpoint Method

To calculate elasticity, we will use the average percentage change in both quantity and price. This is called the **midpoint method for elasticity** and is represented by the following equations:

$$\text{percent change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1) \div 2} \times 100$$

$$\text{percent change in price} = \frac{P_2 - P_1}{(P_2 + P_1) \div 2} \times 100$$

The advantage of the midpoint method is that one obtains the same elasticity between two price points whether there is a price increase or decrease. This is because the formula uses the same base for both cases.

Calculating the Price Elasticity of Demand

Let's calculate the elasticity from points B to A and from points G to H, shown in Figure 1, below.

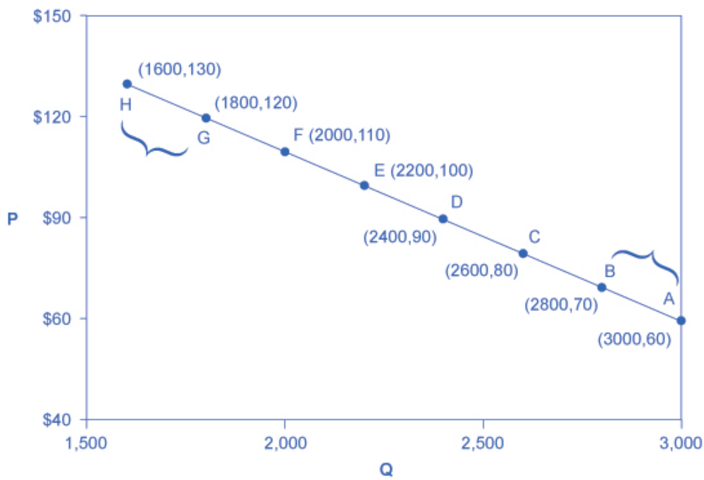


Figure 1. Calculating the Price Elasticity of Demand*. The price elasticity of demand is calculated as the percentage change in quantity divided by the percentage change in price.*

Elasticity from Point B to Point A

Step 1. We know that

$$\text{Price Elasticity of Demand} = \frac{\text{percent change in quantity}}{\text{percent change in price}}$$

Step 2. From the midpoint formula we know that

$$\text{percent change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1) \div 2} \times 100$$

$$\text{percent change in price} = \frac{P_2 - P_1}{(P_2 + P_1) \div 2} \times 100$$

Step 3. We can use the values provided in the figure (as price decreases from \$70 at point B to \$60 at point A) in each equation:

$$\text{percent change in quantity} = \frac{3,000 - 2,800}{(3,000 + 2,800) \div 2} \times 100 = \frac{200}{2,900} \times 100 = 6.9$$

$$\text{percent change in price} = \frac{60 - 70}{(60 + 70) \div 2} \times 100 = \frac{-10}{65} \times 100 = -15.4$$

Step 4. Then, those values can be used to determine the price elasticity of demand:

$$\text{Price Elasticity of Demand} = \frac{6.9 \text{ percent}}{-15.5 \text{ percent}} = -0.45$$

The elasticity of demand between these two points is 0.45, which is an amount smaller than 1. That means that the demand in this interval is inelastic.

Price elasticities of demand are *always* negative, since price and quantity demanded always move in opposite directions (on the demand curve). As you'll recall, according to the law of demand, price and quantity demanded are inversely related. By convention, we always talk about elasticities as positive numbers, however. So, mathematically, we take the absolute value of the result. For example, -0.45 would interpreted as 0.45.

This means that, along the demand curve between points B and A, if the price changes by 1%, the quantity demanded will change by 0.45%. A change in the price will result in a smaller percentage change in the quantity demanded. For example, a 10% *increase* in the price will result in only a 4.5% *decrease* in quantity demanded. A 10% *decrease* in the price will result in only a 4.5% *increase* in the quantity demanded. Price elasticities of demand are negative numbers indicating that the demand curve is downward sloping, but they're read as absolute values.

Elasticity from Point G to Point H

Calculate the price elasticity of demand using the data in Figure 1 for an increase in price from G to H. Does the elasticity increase or decrease as we move up the demand curve?

Step 1. We know that

$$\text{Price Elasticity of Demand} = \frac{\text{percent change in quantity}}{\text{percent change in price}}$$

Step 2. From the midpoint formula we know that

$$\text{percent change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1) \div 2} \times 100$$
$$\text{percent change in price} = \frac{P_2 - P_1}{(P_2 + P_1) \div 2} \times 100$$

Step 3. We can use the values provided in the figure in each equation:

$$\text{percent change in quantity} = \frac{1,600 - 1,800}{(1,600 + 1,800) \div 2} \times 100 = \frac{-200}{1,700} \times 100 = -11.76$$
$$\text{percent change in price} = \frac{130 - 120}{(130 + 120) \div 2} \times 100 = \frac{10}{125} \times 100 = 8.0$$

Step 4. Then, those values can be used to determine the price elasticity of demand:

$$\text{Price Elasticity of Demand} = \frac{\text{percent change in quantity}}{\text{percent change in price}} = \frac{-11.76}{8} = 1.45$$

The elasticity of demand from G to H is 1.47. The magnitude of the elasticity has increased (in absolute value) as we moved up along the demand curve from points A to B. Recall that the elasticity between those two points is 0.45. Demand is inelastic between points A and B and elastic between points G and H. This shows us that price elasticity of demand changes at different points along a *straight-line demand curve*.

Let's pause and think about why the elasticity is different over different parts of the demand curve. When price elasticity of demand is greater (as between points G and H), it means that there is a larger impact on demand as price changes. That is, when the price is higher, buyers are more sensitive to additional price increases. Logically, that makes sense.

Calculating the Price Elasticity of Supply

Let's start with the definition:

> **Price elasticity of supply** is the percentage change in the quantity of a good or service supplied divided by the percentage change in the price.

Elasticity from Point A to Point B

Assume that an apartment rents for $650 per month and at that price 10,000 units are offered for rent, as shown in Figure 2, below. When the price increases to $700 per month, 13,000 units are offered for rent. By what percentage does apartment supply increase? What is the price sensitivity?

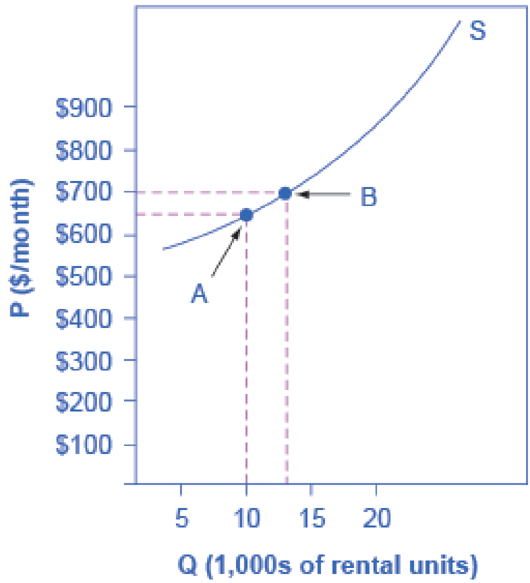


Figure 2. Price Elasticity of Supply*. The price elasticity of supply is calculated as the percentage change in quantity divided by the percentage change in price.*

Step 1. We know that

$$\text{Price Elasticity of Supply} = \frac{\text{percent change in quantity}}{\text{percent change in price}}$$

Step 2. From the midpoint formula we know that

$$\text{percent change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1) \div 2} \times 100$$

$$\text{percent change in price} = \frac{P_2 - P_1}{(P_2 + P_1) \div 2} \times 100$$

Step 3. We can use the values provided in the figure in each equation:

$$\text{percent change in quantity} = \frac{13,000 - 10,000}{(13,000 + 10,000) \div 2} \times 100 = \frac{3,000}{11,500} \times 100 = 26.1$$

$$\text{percent change in price} = \frac{750 - 600}{(750 + 600) \div 2} \times 100 = \frac{50}{675} \times 100 = 7.4$$

Step 4. Then, those values can be used to determine the price elasticity of demand:

$$\text{Price Elasticity of Supply} = \frac{26.1 \text{ percent}}{7.4 \text{ percent}} = 3.53$$

Again, as with the elasticity of demand, the elasticity of supply is not followed by any units. Elasticity is a ratio of one percentage change to another percentage change—nothing more—and is read as an absolute value. In this case, a 1% rise in price causes an increase in quantity supplied of 3.5%. Since 3.5 is greater than 1, this means that the percentage change in quantity supplied will be greater than a 1% price change. If you're starting to wonder if the concept of slope fits into this calculation, read the following example.

Elasticity Is Not Slope

It's a common mistake to confuse the slope of either the supply or demand curve with its elasticity. The slope is the rate of change in units along the curve, or the rise/run (change in y over the change in x). For example, in Figure 1, for each point shown on the demand curve, price drops by $10 and the number of units demanded increases by 200. So the slope is –10/200 along the entire demand curve, and it doesn't change. The price elasticity, however, changes along the curve. Elasticity between points B and A was 0.45 and increased to 1.47 between points G and H. Elasticity is the *percentage* change—which is a different calculation from the slope, and it has a different meaning.

When we are at the upper end of a demand curve, where price is high and the quantity demanded is low, a small change in the quantity demanded—even by, say, one unit—is pretty big in percentage terms. A change in price of, say, a dollar, is going to be much less important in percentage terms than it will be at the bottom of the demand curve. Likewise, at the bottom of the demand curve, that one unit change when the quantity demanded is high will be small as a percentage. So, at one end of the demand curve, where we have a large percentage change in quantity demanded over a small percentage change in price, the elasticity value will be high—demand will be relatively elastic. Even with the same change in the price and the same change in the quantity demanded, at the other end of the demand curve the quantity is much higher, and the price is much lower, so the percentage change in quantity demanded is smaller and the percentage change in price is much higher. See Figure 3, below:

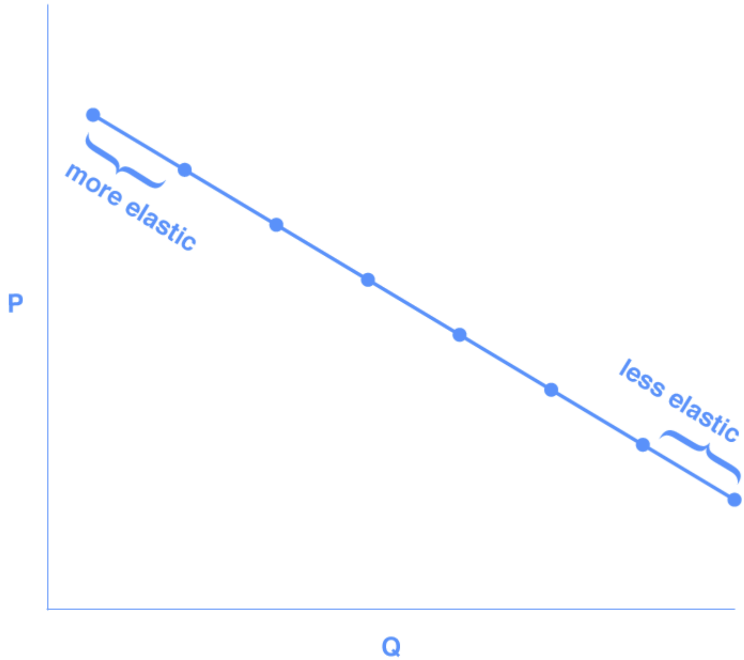


Figure 3. Elasticity Changes Along the Demand Curve

At the bottom of the curve we have a small numerator over a large denominator, so the elasticity measure will be much lower, or inelastic. As we move along the demand curve, the values for quantity and price go up or down, depending on which way we are moving, so the percentages for, say, a $1 difference in price or a one-unit difference in quantity, will change as well, which means the ratios of those percentages will change, too.

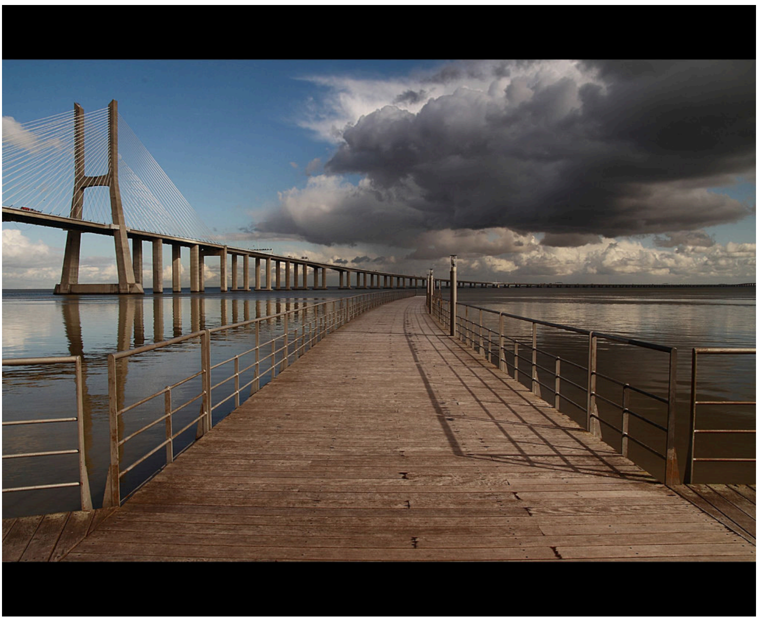
86. Reading: Three Categories of Elasticity

It's helpful to divide elasticities into three categories: elastic, inelastic, and unitary. An **elastic demand** or **elastic supply** is one in which the elasticity is greater than 1, indicating a high responsiveness to changes in price. Elasticities that are less than 1 indicate low responsiveness to price changes and correspond to **inelastic demand** or **inelastic supply**. **Unitary elasticities** indicate proportional responsiveness of either demand or supply. In other words, the change in demand or supply is equal to the change in price, and the elasticities equal 1. These ranges are summarized in Table 1, below.

Table 1.Three Categories of Elasticity: Elastic, Inelastic, and Unitary

If . . .	**Then . . .**	**And It's Called . . .**
% change in quantity > % change in price	% change in quantity % change in price > 1	Elastic
% change in quantity = % change in price	% change in quantity % change in price = 1	Unitary
% change in quantity < % change in price	% change in quantity % change in price < 1	Inelastic

87. Reading: Polar Cases of Elasticity



Polar Cases of Elasticity

There are two extreme cases of elasticity: when elasticity equals zero and when it's infinite. We will describe each case.

Infinite elasticity or **perfect elasticity** refers to the extreme case in which either the quantity demanded (Qd) or supplied (Qs) changes by an infinite amount in response to any change in price at all. In both cases, the supply curve and the demand curve are horizontal, as shown in Figure 1, below.

Perfectly elastic supply is unrealistic; however, the curve can be explained using a little imagination. If supply is perfectly elastic, it means that any change in price will result in an infinite amount of change in quantity. Suppose that you baked delicious cookies and your costs, including inputs and time, were $3 per cookie. At $3, you would be willing to sell as many cookies as you could. You would not sell a single cookie if the price were any lower than $3, and if price were above $3, you would sell an infinite amount. In summary, your supply curve would be perfectly elastic at a price of $3, and any change in price would result in a change in quantity supplied to infinity or zero, depending on whether price increased or decreased, respectively.

Similarly, **perfectly elastic demand** is an extreme example. Perfect elastic demand means that quantity demanded will increase to infinity when the price decreases, and quantity demanded will decrease to zero when price increases. When consumers are extremely sensitive to changes in price, you can think about perfectly elastic demand as "all or nothing." For example, if the price of cruises to the Caribbean decreased, everyone would buy tickets (i.e., quantity demanded would increase to infinity), and if the price of cruises to the Caribbean increased, not a single person would be on the boat (i.e., quantity demanded would decrease to zero).

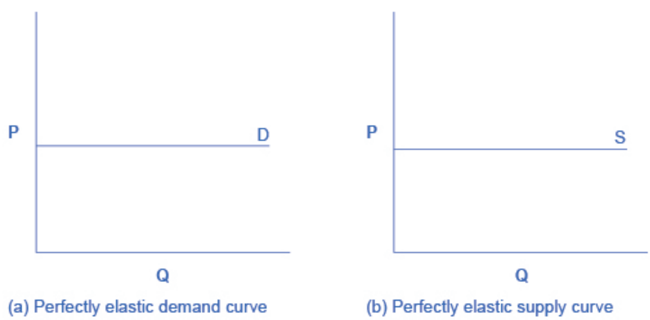


Figure 1. Infinite Elasticity*. The horizontal lines show that an infinite quantity will be demanded or supplied at a specific price. This illustrates the cases of a perfectly (or infinitely) elastic demand curve and supply curve. The quantity supplied or demanded is extremely responsive to price changes, moving from zero for prices close to P to infinite when prices reach P.*

Zero elasticity or **perfect inelasticity**, as depicted in Figure 2, refers to the extreme case in which a percentage change in price, no matter how large, results in zero change in quantity supplied or demanded.

While a **perfectly inelastic supply** is an extreme example, goods with limited supply of inputs are likely to feature highly inelastic supply curves. Consider housing in prime locations such as apartments facing Central Park in New York City or beachfront property in Southern California. If housing prices increase for beachfront property in Southern California, there is a fixed amount of land, and only so many houses can be squeezed in along the beach. If housing prices decrease for Central Park–facing apartments, sellers are not going to bulldoze the buildings. Perfectly inelastic supply means that quantity supplied remains the same when price increases or decreases. Sellers are completely unresponsive to changes in price.

Similarly, while **perfectly inelastic demand** is an extreme case, necessities with no close substitutes are likely to have highly inelastic demand curves. This is the case with life-saving prescription drugs, for example. Consider a person with kidney failure who needs insulin to stay alive. A specific quantity of insulin is prescribed to the patient. If the price of insulin decreases, the patient can't stock up and save it for the future. If the price of insulin increases, the patient will continue to purchase the same quantity needed to stay alive. Perfectly inelastic demand means that quantity demanded remains the same when price increases or decreases. Consumers are completely unresponsive to changes in price.

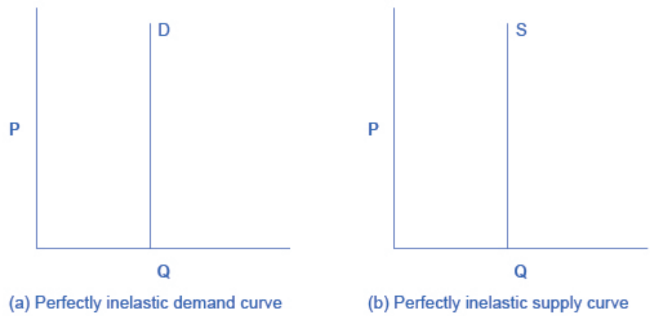


Figure 2. Zero Elasticity*. The vertical supply curve and vertical demand curve show that there will be zero percentage change in quantity (a) supplied or (b) demanded, regardless of the price. This illustrates the case of zero elasticity (or perfect inelasticity). The quantity supplied or demanded is not responsive to price changes.*

Self Check: Calculating Price Elasticity

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the four Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1565

88. Reading: Other Types of Elasticity

S'more ingredients: negative or positive cross-price elasticities of demand?

Introduction

The basic idea of elasticity—how a percentage change in one variable causes a percentage change in another variable—does not just apply to the responsiveness of supply and demand to changes in the price of a product. Recall that quantity demanded (Qd) depends on income, tastes and preferences, population, expectations about future prices, and the prices of related goods. Similarly, quantity supplied (Qs) depends on the cost of production, changes in weather (and natural conditions), new technologies, and government policies. Elasticity can be measured for any determinant of supply and demand, not just the price.

Income Elasticity of Demand

The **income elasticity of demand** is the percentage change in quantity demanded divided by the percentage change in income, as follows:

$$\text{income elasticity of demand} = \frac{\text{percent change in quantity demanded}}{\text{percent change in income}}$$

For most products, most of the time, the income elasticity of demand is positive: that is, a rise in income will cause an increase in the quantity demanded. This pattern is common enough that these goods are referred to as **normal goods**. However, for a few goods, an increase in income means that one might purchase less of the good; for example, those with a higher income might buy fewer hamburgers, because they are buying more steak instead, or those with a higher income might buy less cheap wine and more imported beer. When the income elasticity of demand is negative, the good is called an **inferior good**. The concepts of normal and inferior goods were introduced in the Supply and Demand module. A higher level of income for a normal good causes a demand curve to shift to the right for a normal good, which means that the income elasticity of demand is positive. How far the demand shifts depends on the income elasticity of demand. A higher income elasticity means a larger shift. However, for an inferior good—that is, when the income elasticity of demand is negative—a higher level of income would cause the demand curve for that good to shift to the left. Again, how much it shifts depends on how large the (negative) income elasticity is.

Cross-Price Elasticity of Demand

A change in the price of one good can shift the quantity demanded for another good. If the two goods are complements, like bread and peanut butter, then a drop in the price of one good will lead to an increase in the quantity demanded of the other good. However, if the two goods are substitutes, like plane tickets and train tickets, then a drop in the price of one good will cause people to substitute toward that good, and to reduce consumption of the other good. Cheaper plane tickets lead to fewer train tickets, and vice versa. The **cross-price elasticity of demand** puts some meat on the bones of these ideas. The term "cross-price" refers to the idea that the price of one good is affecting the quantity demanded of a different good. Specifically, the cross-price elasticity of demand is the percentage change in the quantity of good A that is demanded as a result of a percentage change in the price of good B, as follows:

$$\text{cross-price elasticity of demand} = \frac{\text{percent change in } Qd \text{ of good } A}{\text{percent change in price of good } B}$$

Substitute goods have positive cross-price elasticities of demand: if good A is a substitute for good B, like coffee and tea, then a higher price for B will mean a greater quantity of A consumed. Complement goods have negative cross-price elasticities: if good A is a complement for good B, like coffee and sugar, then a higher price for B will mean a lower quantity of A consumed.

Elasticity in Labor and Financial Capital Markets

The concept of elasticity applies to any market, not just markets for goods and services. In the labor market, for example, the **wage elasticity of labor supply**—that is, the percentage change in hours worked divided by the percentage change in wages—will determine the shape of the labor supply curve. The formula is as follows:

$$\text{elasticity of labor supply} = \frac{\text{percent change in quantity of labor supplied}}{\text{percent change in wage}}$$

The wage elasticity of labor supply for teenage workers is generally thought to be fairly elastic: That is, a certain percentage change in wages will lead to a larger percentage change in the quantity of hours worked. Conversely, the wage elasticity of labor supply for adult workers in their thirties and forties is thought to be fairly inelastic. When wages move up or down by a certain percentage amount, the quantity of hours that adults in their prime earning years are willing to supply changes but by a lesser percentage amount. In markets for financial capital, the **elasticity of savings**—that is, the percentage change in the quantity of savings divided by the percentage change in interest rates—will describe the shape of the supply curve for financial capital, as follows:

$$\text{elasticity of savings} = \frac{\text{percent change in quantity of financial savings}}{\text{percent change in interest rate}}$$

Sometimes laws are proposed that seek to increase the quantity of savings by offering tax breaks so that the return on savings is higher. Such a policy will increase the quantity if the supply curve for financial capital is elastic, because then a given percentage increase in the return to savings will cause a higher percentage increase in the quantity of savings. However, if the supply curve for financial capital is highly inelastic, then a percentage increase in the return to savings will cause only a small increase in the quantity of savings. The evidence on the supply curve of financial capital is controversial but, at least in the short run, the elasticity of savings with respect to the interest rate appears fairly inelastic.

Expanding the Concept of Elasticity

The elasticity concept does not even need to relate to a typical supply or demand curve at all. For example, imagine that you are studying whether the Internal Revenue Service should spend more money on auditing tax returns. The question can be framed in terms of the elasticity of tax collections with respect to spending on tax enforcement; that is, what is the percentage change in tax collections derived from a percentage change in spending on tax enforcement? With all of the elasticity concepts that have just been described, some of which are listed in Table 1, the possibility of confusion arises. When you hear the phrases “elasticity of demand” or “elasticity of supply,” they refer to the elasticity with respect to price. Sometimes, either to be extremely clear or because a wide variety of elasticities is being discussed, the elasticity of demand or the demand elasticity will be called the price elasticity of demand or the “elasticity of demand with respect to price.” Similarly, elasticity of supply or the supply elasticity is sometimes called, to avoid any possibility of confusion, the price elasticity of supply or “the elasticity of supply with respect to price.” But in whatever context elasticity is invoked, the idea always refers to percentage change in one variable, almost always a price or money variable, and how it causes a percentage change in another variable, typically a quantity variable of some kind.

Table 1. Formulas for Calculating Elasticity

Elasticity Type	**Formula**
Income elasticity of demand	=%change in Qd / % change in income
Cross-price elasticity of demand	=% change in Qd of good A / % change in price of good B
Wage elasticity of labor supply	=% change in quantity of labor supplied / % change in wage
Wage elasticity of labor demand	=% change in quantity of labor demanded / % change in wage
Interest rate elasticity of savings	=% change in quantity of savings / % change in interest rate
Interest rate elasticity of borrowing	=% change in quantity of borrowing / % change in interest rate

89. Outcome: Other Elasticities

What you'll learn to do: explain and calculate other elasticities using common economic variables

Remember, we elasticity measures the responsiveness of one variable to changes in another variable. We have focused on how a change in price can impact other variables. Elasticity doesn't apply only to price, however. It can describe anything that affects demand/supply. For example, when consumer income varies, it can have an impact on demand. When we consider that impact, we are measuring the responsiveness of one variable (demand) to changes in another variable (consumer income). This is called the income elasticity of demand.

Likewise, if two goods are complements or substitutes, a change in demand for one can have an impact on the demand for the other. This is known as cross-price elasticity of demand. In this section, we'll elaborate on the idea of elasticity to see how it applies to other economic variables.

The specific things you'll learn in this section include the following:

- Explain and calculate elasticity of income and labor
- Explain and calculate cross-price elasticity of demand

Learning Activities

The learning activities for this section include the following:

- Reading: Other Types of Elasticity
- Worked Example: Cross-Price Elasticity of Demand
- Self Check: Other Elasticities

90. Worked Example: Cross-Price Elasticity of Demand

Calculating Cross-Price Elasticity of Demand

This worked example asks you to compute two types of demand elasticities and then to draw conclusions from the results. The initial price and quantity of widgets demanded is (P1 = 12, Q1 = 8). The subsequent price and quantity is (P2 = 9, Q2 = 10). This is all the information needed to compute the price elasticity of demand.

The price elasticity of demand is defined as follows:

$$\text{Price Elasticity of Demand} = \frac{\text{percent change in quantity}}{\text{percent change in price}}$$

From the midpoint formula, we know that:

$$\text{percent change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1) \div 2} \times 100 = \frac{10 - 8}{(10 + 8) \div 2} \times 100 = \frac{2}{9} \times 100 = 22.2$$

And:

$$\text{percent change in price} = \frac{P_2 - P_1}{(P_2 + P_1) \div 2} \times 100 = \frac{9 - 12}{(9 + 12) \div 2} \times 100 = \frac{-3}{10.5} \times 100 = -28.6$$

Therefore:

$$\text{Price Elasticity of Demand} = \frac{22.2 \text{ percent}}{-28.6 \text{ percent}} = -0.77$$

Since the elasticity is less than 1 (in absolute value), we can say that the price elasticity of demand for widgets is in the inelastic range.

The cross-price elasticity of demand is computed similarly:

$$\text{Cross-Price Elasticity of Demand} = \frac{\text{percent change in quantity of sprockets demanded}}{\text{percent change in price of widgets}}$$

The initial quantity of sprockets demanded is 9 and the subsequent quantity demanded is 10 (Q1 = 9, Q2 = 10).

Using the midpoint formula, we can calculate the percent change in the quantity of sprockets demanded:

$$\text{percent change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1) \div 2} \times 100 = \frac{10 - 9}{(10 + 9) \div 2} \times 100 = \frac{1}{9.5} \times 100 = 10.5$$

The percent change in the quantity of sprockets demanded is 10.5%.

The percent change in the price of widgets is the same as above, or -28.6%.

Therefore:

$$\text{Cross-Price Elasticity of Demand} = \frac{10.5 \text{ percent}}{-28.6 \text{ percent}} = -0.37$$

Because the cross-price elasticity is negative, we can conclude that widgets and sprockets are complementary goods. Intuitively, when the price of widgets goes down, consumers purchase more widgets. Because they're purchasing more widgets, they purchase more sprockets.

91. Self Check: Other Elasticities

Check Your Understanding

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1566

92. Outcome: Price Elasticity and Total Revenue

What you'll learn to do: explain the relationship between a firm's price elasticity of demand and total revenue

Price elasticity of demand describes how changes in the price for goods and the demand for those same goods relate. As those two variables interact, they can have an impact on a firm's total revenue. Revenue is the amount of money a firm brings in from sales—i.e., the total number of units sold multiplied by the price per unit. Therefore, as the price or the quantity sold changes, those changes have a direct impact on revenue.

Businesses seek to maximize their profits, and price is one tool they have at their disposal to influence demand (and therefore sales). Picking the right price is tricky, though. What happens with a price increase? Will customers buy only a little less, such that the price increase raises revenues, or will they buy a lot less, such that the price increase lowers revenues? Might the company earn more if it lowers prices, or will that just lead to lower revenue per unit without stimulating new demand? These are critical questions for every business.

In this section, you'll learn more about how firms think about the impact of price elasticities on revenue.

The specific things you'll learn in this section include the following:

- Explain the interaction between price and revenue, given elastic demand
- Explain the interaction between price and revenue, given

inelastic demand
- Explain how changes in production costs affect price

Learning Activities

The learning activities for this section include the following:

- Reading: Elasticity and Total Revenue
- Reading: Elasticity, Costs, and Customers
- Self Check: Price Elasticity and Total Revenue

93. Reading: Elasticity and Total Revenue

Total Revenue and Elasticity of Demand

Studying elasticities is useful for a number of reasons, pricing being the most important. The key consideration when thinking about maximizing revenue is the price elasticity of demand. **Total revenue** is the price of an item multiplied by the number of units sold: TR = P x Qd. When a firm considers a price increase or decrease, there are three possibilities, which are laid out in Table 1, below.

Table 1. Price Elasticity of Demand

If demand is . . .	Then . . .	Therefore . . .
Elastic	% change in Qd is greater than % change in P	A given % rise in P will be more than offset by a larger % fall in Q so that total revenue (P times Q) falls.
Unitary	% change in Qd is equal to % change in P	A given % rise in P will be exactly offset by an equal % fall in Q so that total revenue (P times Q) is unchanged.
Inelastic	% change in Qd is less than % change in P	A given % rise in P will cause a smaller % fall in Q so that total revenue (P times Q) rises.

If demand is elastic at a given price level, then the company should cut its price, because the percentage drop in price will result in an even larger percentage increase in the quantity sold—thus raising total revenue. However, if demand is inelastic at the original quantity level, then the company should raise its prices, because the percentage increase in price will result in a smaller percentage decrease in the quantity sold—and total revenue will rise.

Let's explore some specific examples. In both cases we will answer the following questions:

1. How much of an impact do we think a price change will have on demand?
2. How would we calculate the elasticity, and does it confirm our assumption?
3. What impact does the elasticity have on total revenue?

Example 1: The Student Parking Permit

How elastic is the demand for student parking passes at your institution? The answer to that question likely varies based on the profile of your institution, but we are going to explore a particular example. Let's consider a community college campus where all of the students commute to class. Required courses are spread throughout the day and the evening, and most of the classes require classroom attendance (rather than online participation). There is a reasonable public transportation system with busses coming to and leaving campus from several lines, but the majority of students drive to campus. A student parking permit costs $40 per term. As the parking lots become increasingly congested, the college considers raising the price of the parking passes in hopes that it will encourage more students to carpool or to take the bus.

If the college increases the price of a parking permit from $40 to $48, will fewer students buy parking permits?

If you think that the change in price will cause many students to decide not to buy a permit, then you are suggesting that the demand is elastic—the students are quite sensitive to price changes. If you think that the change in price will not impact student permit purchases much, then you are suggesting that the demand is inelastic—student demand for permits is insensitive to price changes.

In this case, we can all argue that students are very sensitive to increases in costs *in general*, but the determining factor in their demand for parking permits is more likely to be the quality of alternative solutions. If the bus service does not allow students to travel between home, school, and work in a reasonable amount of time, many students will resort to buying a parking permit, even at the higher price. Because students don't generally have extra money, they may grumble about a price increase, but many will still have to pay.

Let's add some numbers and test our thinking. The college implements the proposed increase of $8, taking the new price to $48. Last year the college sold 12,800 student parking passes. This year, at the new price, the college sells 11,520 parking passes.

$$\text{percent change in quantity} = \frac{11,520 - 12,800}{(11,520 + 12,800) \div 2} \times 100 = \frac{-1280}{12160} \times 100 = -10.53$$

$$\text{percent change in price} = \frac{48 - 40}{(48 + 40) \div 2} \times 100 = \frac{8}{44} \times 100 = 18.18$$

$$\text{Price Elasticity of Demand} = \frac{-10.53 \text{ percent}}{18.18 \text{ percent}} = -.58$$

First, looking only at the percent change in quantity and the percent change in price we know that an 18% change in price will resulted in an 11% change in demand. In other words, a large change in price created a comparatively smaller change in demand. We can also see that the elasticity is 0.58. When the absolute value of the price elasticity is < 1, the demand is inelastic. In this example, student demand for parking permits is inelastic.

What impact does the price change have on the college and their goals for students? First, there are 1,280 fewer cars taking up parking places. If all of those students are using alternative transportation to get to school and this change has relieved parking-capacity issues, then the college may have achieved its goals. However, there's more to the story: the price change also has an effect on the college's revenue, as we can see below:

Year 1: 12,800 parking permits sold x $40 per permit = $512,000

Year 2: 11,520 parking permits sold x $48 per permit = $552,960

The college earned an additional $40,960 in revenue. Perhaps this can be used to expand parking or address other student transportation issues.

In this case, student demand for parking permits is inelastic. A significant change in price leads to a comparatively smaller change in demand. The result is lower sales of parking passes but more revenue.

Note: If you attend an institution that offers courses completely or largely online, the price elasticity for parking permits might be perfectly inelastic. Even if the institution gave away parking permits, students might not want them.

Example 2: Helen's Cookies

Have you been at the counter of a convenience store and seen cookies for sale on the counter? In this example we are going to consider a baker, Helen, who bakes these cookies and sells them $2 each. The cookies are sold in a convenience store, which has several options on the counter that customers can choose as a last-minute impulse buy. All of the impulse items range between $1 and $2 in price. In order to raise revenue, Helen decides to raise her price to $2.20.

If Helen increases the cookie price from $2.00 to $2.20—a 10% increase—will fewer customers buy cookies?

If you think that the change in price will cause many buyers to forego a cookie, then you are suggesting that the demand is elastic, or that the buyers are sensitive to price changes. If you think that the change in price will not impact sales much, then you are suggesting that the demand for cookies is inelastic, or insensitive to price changes.

Let's assume that this price change does impact customer behavior. Many customers choose a $1 chocolate bar or a $1.50 doughnut over the cookie, or they simply resist the temptation of the cookie at the higher price. Before we do any math, this assumption suggests that the demand for cookies is elastic.

Adding in the numbers, we find that Helen's weekly sales drop from 200 cookies to 150 cookies. This is a 25% change in demand on account of a 10% price increase. We immediately see that the change in demand is greater than the change in price. That means that demand is elastic. Let's do the math.

$$\text{percent change in quantity} = \frac{150-200}{(150+200)\div 2} \times 100 = \frac{-50}{175} \times 100 = -28.75$$

$$\text{percent change in price} = \frac{2.20-2.00}{(2.00+2.20)\div 2} \times 100 = \frac{.20}{2.10} \times 100 = 9.52$$

$$\text{Price Elasticity of Demand} = \frac{-28.75 \text{ percent}}{9.52 \text{ percent}} = -3$$

When the absolute value of the price elasticity is > 1, the demand is elastic. In this example, the demand for cookies is elastic.

What impact does this have on Helen's objective to increase revenue? It's not pretty.

Price 1: 200 cookies sold x $2.00 per cookie = $400

Price 2: 150 cookies sold x $2.20 = $330

She is earning less revenue because of the price change. What should Helen do next? She has learned that a small change in price leads to a large change in demand. What if she lowered the price slightly from her original $2.00 price? If the pattern holds, then a small reduction in price will lead to a large increase in sales. That would give her a much more favorable result.

94. Reading: Elasticity, Costs, and Customers

Customers and Changing Costs

We can see that understanding elasticity helps a firm set a price that maximizes total revenue. What happens if the firm's production costs change, though? And what is the impact on customers?

Most businesses are continually trying to figure out ways to produce at a lower cost, as one path to earning higher profits. It's a challenge to do this, though, when the price of a key input over which a firm has no control rises. For example, many chemical companies use petroleum as a key input, but they have no control over the world market price for crude oil. Coffee shops use coffee as a key input, but they have no control over the world market price of coffee. If the cost of a key input rises, can the firm pass along those higher costs to consumers in the form of higher prices?

Conversely, if new and less expensive ways of producing are invented, can the firm keep the benefits in the form of higher profits, or will the market pressure them to pass along the gains to consumers in the form of lower prices? The price elasticity of demand plays a key role in answering these questions.

Imagine that, as a consumer of legal pharmaceutical products, you read a news story about a technological breakthrough in the production of aspirin: Now every aspirin factory can make aspirin more cheaply than it did before.

What does this discovery mean to the firm? Figure 1, below, illustrates two possibilities. In Figure 1 (a), the demand curve is drawn as highly inelastic. In this case, a technological breakthrough that shifts supply to the right, from S0 to S1, so that the equilibrium shifts from E0 to E1, creates a substantially lower price for the product with relatively little impact on the quantity sold. In Figure 1 (b), the demand curve is drawn as highly elastic. In this case, the technological breakthrough leads to a much greater quantity being sold in the market at very close to the original price. Consumers benefit more, in general, when the demand curve is more inelastic because the shift in the supply results in a much lower price for consumers.

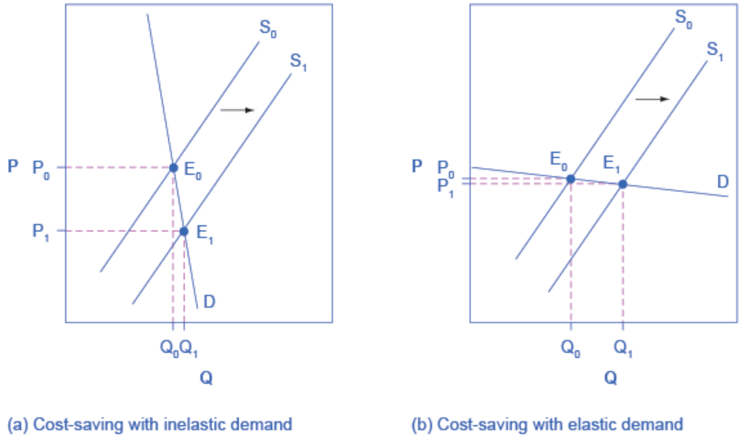


Figure 1. Passing along Cost Savings to Consumers*. Cost-saving gains cause supply to shift out to the right from S0 to S1; that is, at any given price, firms will be willing to supply a greater quantity. If demand is inelastic, as in (a), the result of this cost-saving technological improvement will be substantially lower prices. If demand is elastic, as in (b), the result will be only slightly lower prices. Consumers benefit in either case, from a greater quantity at a lower price, but the benefit is greater when demand is inelastic, as in (a).*

Producers of aspirin may find themselves in a nasty bind here. The situation shown in Figure 1, with extremely inelastic demand, means that a new invention may cause the price to drop dramatically while quantity changes little. As a result, the new production technology can lead to a drop in the revenue that firms earn from sales of aspirin. However, if strong competition exists between producers of aspirin, each producer may have little choice but to search for and implement any breakthrough that allows it to reduce production costs.

After all, if one firm decides not to implement such a cost-saving technology, it can be driven out of business by other firms that do.

Since demand for food is generally inelastic, farmers may often face the situation in Figure 1 (a). That is, a surge in production leads to a severe drop in price that can actually decrease the total revenue received by farmers. Conversely, poor weather or other conditions that cause a terrible year for farm production can sharply raise prices so that the total revenue received increases. The example below discusses how these issues relate to coffee.

Fluctuations in Coffee Prices

Coffee is an international crop. The top five coffee-exporting nations are Brazil, Vietnam, Colombia, Indonesia, and Guatemala. In these nations and others, 20 million families depend on selling coffee beans as their main source of income. These families are exposed to enormous risk, because the world price of coffee bounces up and down. For example, in 1993, the world price of coffee was about 50 cents per pound; in 1995 it was four times as high, at $2 per pound. By 1997 it had fallen by half to $1.00 per pound. In 1998 it leaped back up to $2 per pound. By 2001 it had fallen back to 46 cents a pound; by early 2011 it went back up to about $2.31 per pound. By the end of 2012, the price had fallen back to about $1.31 per pound.

The reason for these price fluctuations stems from a combination of inelastic demand and shifts in supply. The elasticity of coffee demand is only about 0.3; that is, a 10% rise in the price of coffee leads to a decline of about 3% in the quantity of coffee consumed. When a major frost hit the Brazilian coffee crop in 1994, coffee supply shifted to the left with an inelastic demand curve, leading to much higher prices. Conversely, when Vietnam entered the world coffee market as a major producer in the late 1990s, the supply curve shifted out to the right. With a highly inelastic demand curve, coffee prices fell dramatically. This situation is shown in Figure 1 (a), above.

Elasticity also reveals whether firms can pass along higher costs to consumers. Addictive substances tend to fall into this category. For example, the demand for cigarettes is relatively inelastic among regular smokers who are pretty addicted; economic research suggests that increasing the price of cigarettes by 10% leads to about a 3% reduction in the quantity of cigarettes smoked by adults, so the elasticity of demand for cigarettes is 0.3. If society increases taxes on companies that make cigarettes, the result will be, as in Figure 2 (a), that the supply curve shifts from S0 to S1. However, as the equilibrium moves from E0 to E1, these taxes are mainly passed along to consumers in the form of higher prices. These higher taxes on cigarettes will raise tax revenue for the government, but they will not much affect the quantity of smoking.

If the goal is to reduce the quantity of cigarettes demanded, it must be achieved by shifting this inelastic demand back to the left, perhaps with public programs to discourage the use of cigarettes or to help people to quit. For example, antismoking advertising campaigns have shown some ability to reduce smoking. However, if demand for cigarettes were more elastic, as in Figure 2 (b), then an increase in taxes that shifts supply from S0 to S1 and equilibrium from E0 to E1 would reduce the quantity of cigarettes smoked substantially. Youth smoking seems to be more elastic than adult smoking—that is, the quantity of youth smoking will fall by a greater percentage than the quantity of adult smoking in response to a given percentage increase in price. Also, casual smokers and low-income smokers are more responsive to changes in the price of cigarettes (that is, their demand is more elastic).

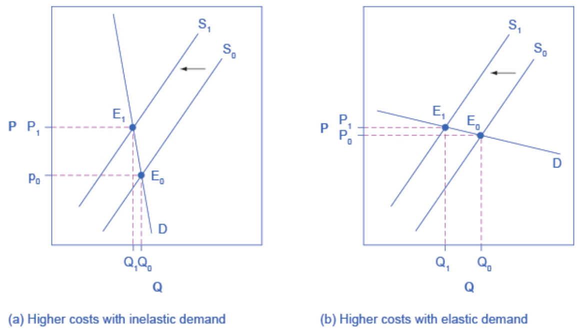


Figure 2. Passing along Higher Costs to Consumers*. Higher costs, like a higher tax on cigarette companies for the example given in the text, lead supply to shift to the left. This shift is identical in (a) and (b). However, in (a), where demand is inelastic, the cost increase can largely be passed along to consumers in the form of higher prices, without much of a decline in equilibrium quantity. In (b), demand is elastic, so the shift in supply results primarily in a lower equilibrium quantity. Consumers suffer in either case, but in (a), they suffer from paying a higher price for the same quantity, while in (b), they suffer from buying a lower quantity (and presumably needing to shift their consumption elsewhere).*

Long-Run vs. Short-Run Impact

Elasticities are often lower in the short run than in the long run. On the demand side of the market, it can sometimes be difficult to change Qd in the short run but easier in the long run. Consumption of energy is a clear example. In the short run, it is not easy for a person to make substantial changes in his or her energy consumption. Maybe you can carpool to work sometimes or adjust your home thermostat by a few degrees if the cost of energy rises, but that's about it. However, in the long-run you can purchase a car that gets more miles to the gallon, choose a job that is closer to where you live, buy more energy-efficient home appliances, or install more insulation in your home.

As a result, the elasticity of demand for energy is somewhat inelastic in the short run, but much more elastic in the long run.

Figure 3 shows an example, based roughly on historical experience, of the responsiveness of Qd to price changes. In 1973, the price of crude oil was $12 per barrel, and total consumption in the U.S. economy was 17 million barrels per day. That year, the nations who were members of the Organization of Petroleum Exporting Countries (OPEC) cut off oil exports to the United States for six months because the Arab members of OPEC disagreed with the U.S. support for Israel. OPEC did not bring exports back to their earlier levels until 1975—a policy that can be interpreted as a shift of the supply curve to the left in the U.S. petroleum market. Figure 3 (a) and (b) show the same original equilibrium point and the same identical shift of a supply curve to the left from S0 to S1.

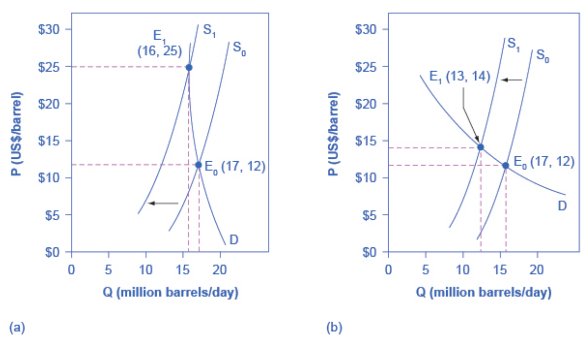


***Figure 3. How a Shift in Supply Can Affect Price or Quantity**. The intersection (E0) between demand curve D and supply curve S0 is the same in both (a) and (b). The shift of supply to the left from S0 to S1 is identical in both (a) and (b). The new equilibrium (E1) has a higher price and a lower quantity than the original equilibrium (E0) in both (a) and (b). However, the shape of the demand curve D is different in (a) and (b). As a result, the shift in supply can result either in a new equilibrium with a much higher price and an only slightly smaller quantity, as in (a), or in a new equilibrium with only a small increase in price and a relatively larger reduction in quantity, as in (b).*

Figure 3 (a) shows inelastic demand for oil in the short run similar to that which existed for the United States in 1973. In Figure 3 (a), the new equilibrium (E1) occurs at a price of $25 per barrel, roughly double the price before the OPEC shock, and an equilibrium quantity of 16 million barrels per day. Figure 3 (b) shows what the outcome would have been if the U.S. demand for oil had been more elastic, a result more likely over the long term. This alternative equilibrium (E1) would have resulted in a smaller price increase to $14 per barrel and larger reduction in equilibrium quantity to 13 million barrels per day. In 1983, for example, U.S. petroleum consumption was 15.3 million barrels a day, which was lower than in 1973 or 1975. U.S. petroleum consumption was down even though the U.S. economy was about one-fourth larger in 1983 than it had been in 1973. The primary reason for the lower quantity was that higher energy prices spurred conservation efforts, and after a decade of home insulation, more fuel-efficient cars, more efficient appliances and machinery, and other fuel-conserving choices, the demand curve for energy had become more elastic.

On the supply side of markets, producers of goods and services typically find it easier to expand production in the long term of several years rather than in the short run of a few months. After all, in the short run it can be costly or difficult to build a new factory, hire many new workers, or open new stores. But over a few years, all of these are possible.

Indeed, in most markets for goods and services, prices bounce up and down more than quantities in the short run, but quantities often move more than prices in the long run. The underlying reason for this pattern is that supply and demand are often inelastic in the short run, so that shifts in either demand or supply can cause a relatively greater change in prices. But since supply and demand are more elastic in the long run, the long-run movements in prices are more muted, while quantity adjusts more easily in the long run.

Self Check: Price Elasticity and Total Revenue

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/1567

95. Problem Set: Elasticity

Test your understanding of the learning outcomes in this module by working through the following problems. These problems aren't graded, but they give you a chance to practice before taking the quiz.

If you'd like to try a problem again, you can click the link that reads, "Try another version of these questions."

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=117

96. Putting It Together: Elasticity

Summary

The goal of this module was to explain the importance of elasticity. You learned how to:

- Define the concept of elasticity
- Explain the price elasticity of demand and price elasticity of supply, and compute both using the midpoint method
- Explain and calculate other elasticities using common economic variables
- Explain the relationship between a firm's price elasticity of demand and total revenue

Netflix Pricing Revisited

We began this module discussing a price change that Netflix imposed on its customers. Now that we understand price elasticity, we can better evaluate that case. How did the 60 percent price increase end up for Netflix? It was a very bumpy two-year ride. Before the price increase, there were about 24.6 million U.S. subscribers. After the price increase, 810,000 infuriated customers canceled their Netflix subscriptions, dropping the total number of subscribers to 23.79 million. Fast-forward to June 2013, when there were 36 million streaming Netflix subscribers in the United States. This was an increase of 11.4 million subscribers since the price increase—an average per-quarter growth of about 1.6

million. This growth is less than the 2 million-per-quarter increases Netflix experienced in the fourth quarter of 2010 and the first quarter of 2011.

During the first year after the price increase, the firm's stock price (a measure of future expectations for the firm) fell from about $300 per share to just under $54. By June 2013, the stock price had rebounded to about $200 per share—still off by more than one-third from its high, but definitely improving. What happened? Obviously, Netflix understood the law of demand. Company officials reported, when they announced the price increase, that this could result in the loss of about 600,000 existing subscribers. Using the elasticity of demand formula, it is easy to see that they expected an inelastic response:

$$\begin{aligned}
&= \frac{-600{,}000/[(24 \text{ million} + 24.6 \text{ million})/2]}{\$6/[(\$10 + \$16)/2]} \\
&= \frac{-600{,}000/24.3 \text{ million}}{\$6/\$13} \\
&= \frac{-0.025}{0.46} \\
&= 0.05
\end{aligned}$$

In addition, Netflix officials had expected that the price increase would have little impact on attracting new customers. Netflix anticipated adding up to 1.29 million new subscribers in the third quarter of 2011. It is true that this was slower growth than the firm had experienced over the past year—about 2 million per quarter. Why was the estimate of customers leaving so far off? During the fourteen years after Netflix was founded, there was an increase in the number of close, but not perfect, substitutes. Consumers now had choices ranging from Vudu, Amazon Prime, Hulu, and Redbox to retail stores.

Jaime Weinman reported in *Maclean's* that Redbox kiosks are "a five-minute drive or less from 68 percent of Americans, and it seems that many people still find a five-minute drive more convenient than loading up a movie online." It seems that, in 2012, many consumers still preferred a physical DVD disk over streaming video. What missteps did the Netflix management make? In addition to misjudging the elasticity of demand, by failing to account for close substitutes, it seems they may have also misjudged customers' preferences and tastes. Yet, as the population increases, the preference for streaming video may overtake physical DVD disks. Netflix, the target of numerous late-night talk-show jabs and laughs in 2011, may yet have the last laugh.

97. Glossary: Elasticity

constant unitary elasticity when a given percentage change in price leads to an equal percentage change in quantity demanded or supplied

cross-price elasticity of demand the percentage change in the quantity of good A that is demanded as a result of a percentage change in the quantity of good B demanded

elastic demand when the elasticity of demand is greater than 1, indicating a high responsiveness of quantity demanded to changes in price

elasticity the responsiveness of one variable to changes in another variable

elasticity of savings the percentage change in the quantity of savings divided by the percentage change in interest rates

elastic supply when the elasticity of supply is greater than 1, indicating a high responsiveness of quantity supplied to changes in price

growth rate percentage change: the change in quantity divided by the quantity

income elasticity of demand the percentage change in quantity demanded divided by the percentage change in income

inelastic demand when the elasticity of demand is smaller than 1, indicating a low responsiveness of quantity demanded price changes

inelastic supply when the elasticity of supply is smaller than 1, indicating a low responsiveness of quantity supplied to price changes

inferior good a good for which the quantity demanded falls as income rises, and the quantity demanded rises as income falls; income elasticity of demand for an inferior good is negative

infinite elasticity the extremely elastic situation of demand or supply in which the quantity changes by an infinite amount in response to any change in price; *also called "perfect elasticity"*

normal good a good for which the quantity demanded rises as income rises, and the quantity demanded falls as income falls; income elasticity of demand for a normal good is positive

price elasticity the relationship between the percent change in price resulting in a corresponding percentage change in the quantity demanded or supplied

price elasticity of demand percentage change in the quantity of a good or service *demanded* divided by the percentage change in price

price elasticity of supply percentage change in the quantity of a good or service *supplied* divided by the percentage change in price

total revenue the price of an item multiplied by the number of units sold

unitary elasticity when the calculated elasticity is equal to 1, indicating that a change in the price of the good or service results in a proportional change in the quantity demanded or supplied

wage elasticity of labor supply the percentage change in hours worked divided by the percentage change in wages

zero elasticity the highly inelastic case of demand or supply in which a percentage change in price, no matter how large, results in zero change in the quantity; *also called "perfect inelasticity"*

98. Discussion: Price Elasticity of Demand

Think of three goods for which the demand is inelastic with respect to price. Do these goods ever go on sale? Does understanding the relationship between elasticity and total revenue help you understand why some goods go on sale and others don't? Share your thoughts.

PART VI

MODULE: GOVERNMENT ACTION

99. Why It Matters: Government Action

Why evaluate the consequences of government policies in markets?

In the module on Supply and Demand, we defined a free market as one with no government intervention. In this module, we will explore the outcomes, both anticipated and otherwise, when government does intervene in a market.

Economists believe there are a small number of fundamental principles that explain how economic agents respond in different situations. Two of these principles, which we have already been introduced to, are the laws of demand and supply.

"Always Check Your Work" by Don Nunn. CC-BY-SA.

Governments can pass laws affecting market outcomes, but no law can negate these economic principles. Rather, the principles will manifest themselves in sometimes unanticipated ways, which may subvert the intent of the government policy. This is one of the major conclusions of this module.

The three best examples of this are:

- Price floors – a legal minimum price in a market, e.g. the minimum wage;
- Price ceilings – a legal maximum price in a market, e.g. rent controls in certain cities;

- Tax incidence – who ends up paying a tax? For example, if the local government adds a sales tax on restaurant meals, is the tax paid by the diners or does it come out of the restaurant's profits?

Understanding all the effects, both anticipated and unanticipated, of government intervention in a market is critical to determining whether the policy achieves its goal.

As you go through this module, make sure to keep in mind who is a given policy (e.g., a minimum wage) supposed to help? Only then can you evaluate whether the policy is a good one or not. Check out the following video about minimum wage:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=122#oembed-1

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=122

LEARNING OUTCOMES

- Analyze the consequences of the government setting a binding price ceiling

- Analyze the consequences of the government setting a binding price floor
- Explain the outcome of a binding price ceiling or price floor on the price and quantity of a product sold
- Determine the incidence of a tax using the concepts of consumer and producer surplus
- Define progressive, proportional, and regressive taxes

100. Outcome: Price Ceilings

What you'll learn to do: analyze the consequences of the government setting a binding price ceiling

In this outcome, we will learn what happens when prices are held below a certain level. Governments typically set a price ceiling to protect consumers by making necessary products affordable, but in this section you'll see how this sometimes backfires by creating a market shortage or other unintended consequences.

The specific things you'll learn to do in this section include:

- Identify the market's equilibrium price and quantity under a price ceiling
- Compute and graph the market shortage resulting from a price ceiling

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Price Ceilings
- Self Check: Price Ceilings

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

101. Reading: Price Ceilings

Demand and Supply Model

Controversy sometimes surrounds the prices and quantities established by demand and supply, especially for products that are considered necessities. In some cases, discontent over prices turns into public pressure on politicians, who may then pass legislation to prevent a certain price from climbing "too high" or falling "too low."

The demand and supply model shows how people and firms will react to the incentives provided by these laws to control prices, in ways that will often lead to undesirable consequences. Alternative policy tools can often achieve the desired goals of price control laws, while avoiding at least some of their costs and tradeoffs.

Price Ceilings

Laws that government enacts to regulate prices are called *Price Controls*. Price controls come in two flavors. A *price ceiling* keeps a price from rising above a certain level (the "ceiling"), while a *price floor* keeps a price from falling below a certain level (the "floor"). This section uses the demand and supply framework to analyze price ceilings. The next section discusses price floors.

In many markets for goods and services, demanders outnumber suppliers. Consumers, who are also potential voters, sometimes unite behind a political proposal to hold down a certain price. In some cities, for example, renters have pressed political leaders to pass rent control laws, a price ceiling that usually works by stating that rents can be raised by only a certain maximum percentage each year.

Rent control becomes a politically hot topic when rents begin to rise rapidly. Everyone needs an affordable place to live. Perhaps a change in tastes makes a certain suburb or town a more popular place to live. Perhaps locally-based businesses expand, bringing higher incomes and more people into the area. Changes of this sort can cause a change in the demand for rental housing, as Figure 3.21 illustrates. The original equilibrium (E_0) lies at the intersection of supply curve S_0 and demand curve D_0, corresponding to an equilibrium price of $500 and an equilibrium quantity of 15,000 units of rental housing. The effect of greater income or a change in tastes is to shift the demand curve for rental housing to the right, as shown by the data in Table 3.7 and the shift from D_0 to D_1 on the graph. In this market, at the new equilibrium E_1, the price of a rental unit would rise to $600 and the equilibrium quantity would increase to 17,000 units.

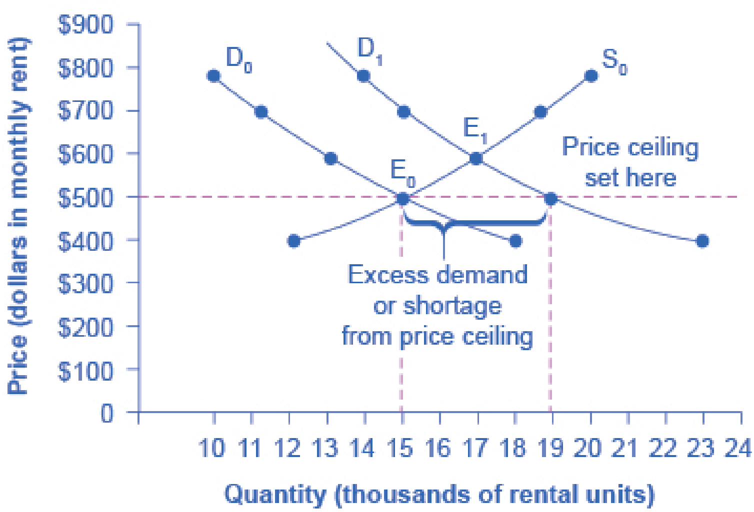


Figure 3.21. A *Price Ceiling Example—Rent Control The original intersection of demand and supply occurs at* E_0*. If demand shifts from* D_0 *to* D_1*, the new equilibrium would be at* E_1*—unless a price ceiling prevents the price from rising. If the price is not permitted to rise, the quantity supplied remains at 15,000. However, after the change in demand, the quantity demanded rises to 19,000, resulting in a shortage.*

Table showing the changes in quantity supplied and quantity demanded at each price.

Table 3.7 Rent Control

Price	Original Quantity Supplied	Original Quantity Demanded	New Quantity Demanded
$400	12,000	18,000	23,000
$500	15,000	15,000	19,000
$600	17,000	13,000	17,000
$700	19,000	11,000	15,000
$800	20,000	10,000	14,000

Suppose that a rent control law is passed to keep the price at the original equilibrium of $500 for a typical apartment. In Figure 3.21, the horizontal line at the price of $500 shows the legally fixed maximum price set by the rent control law. However, the underlying forces that shifted the demand curve to the right are still there. At that price ($500), the quantity supplied remains at the same 15,000 rental units, but the quantity demanded is 19,000 rental units. In other words, the quantity demanded exceeds the quantity supplied, so there is a shortage of rental housing. One of the ironies of price ceilings is that while the price ceiling was intended to help renters, there are actually fewer apartments rented out under the price ceiling (15,000 rental units) than would be the case at the market rent of $600 (17,000 rental units).

Price ceilings do not simply benefit renters at the expense of landlords. Rather, some renters (or potential renters) lose their housing as landlords convert apartments to co-ops and condos. Even when the housing remains in the rental market, landlords tend to spend less on maintenance and on essentials like heating, cooling, hot water, and lighting. The first rule of economics is you do not get something for nothing—everything has an opportunity cost. So if renters get “cheaper” housing than the market requires, they tend to also end up with lower quality housing.

Price ceilings have been proposed for other products. For example, price ceilings to limit what producers can charge have been proposed in recent years for prescription drugs, doctor and hospital fees, the charges made by some automatic teller bank machines, and auto insurance rates. Price ceilings are enacted in an attempt to keep prices low for those who demand the product. But when the market price is not allowed to rise to the equilibrium level, quantity demanded exceeds quantity supplied, and thus a shortage occurs. Those who manage to purchase the product at the lower price given by the price ceiling will benefit, but sellers of the product will suffer, along with those who are not able to purchase the product at all. Quality is also likely to deteriorate.

Self Check: Price Ceilings

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/722

102. Outcome: Price Floors

What you'll learn to do: analyze the consequences of the government setting a binding price floor

In this outcome, we will see what happens when a price floor forces prices above a minimum standard, such as a minimum wage. Watch this video about Edgar the Exploiter to understand how the minimum wage can both help and harm workers.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=125#oembed-1

The specific things you'll learn in this section include:

- Identify the market's equilibrium price and quantity for a price floor
- Compute and graph the market surplus resulting from a price floor
- Explain the outcome of a binding price ceiling or price floor on the price and quantity of a product sold

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Price Floors
- Case in Point: Organic Foods
- Self Check: Price Floors

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

103. Reading: Price Floors

Price Floors

A price floor is the lowest legal price that can be paid in markets for goods and services, labor, or financial capital. Perhaps the best-known example of a price floor is the minimum wage, which is based on the normative view that someone working full-time ought to be able to afford a basic standard of living. The federal minimum wage at the end of 2015 was $7.25 per hour, which yields an income for a single person slightly higher than the poverty line. As the cost of living rises over time, the Congress periodically raises the federal minimum wage.

Image from the NYC Rally To Raise The Minimum Wage, by The All Nite Images. Accessed via Flickr, CC-BY-SA.

Price floors are sometimes called “price supports,” because they support a price by preventing it from falling below a certain level. Around the world, many countries have passed laws to create agricultural price supports. Farm prices and thus farm incomes fluctuate, sometimes widely. So even if, on average, farm incomes are adequate, some years they can be quite low. The purpose of price supports is to prevent these swings.

The most common way price supports work is that the government enters the market and buys up the product, adding to demand to keep prices higher than they otherwise would be. According to Reuters News, the *European Union* (EU) will spend about $60 billion per year, or roughly 38% of the EU budget, on price supports for Europe's farmers from 2014 to 2020.

Figure 3.22 illustrates the effects of a government program that assures a price above the equilibrium by focusing on the market for wheat in Europe. In the absence of government intervention, the price would adjust so that the quantity supplied would equal the quantity demanded at the equilibrium point E_0, with price P_0 and quantity Q_0. However, policies to keep prices high for farmers keeps the price above what would have been the market equilibrium level—the price Pf shown by the dashed horizontal line in the diagram. The result is a quantity supplied in excess of the quantity demanded (Qd). When quantity supplied exceeds quantity demanded, a surplus exists.

The high-income areas of the world, including the United States, Europe, and Japan, are estimated to spend roughly $1 billion per day in supporting their farmers. If the government is willing to purchase the excess supply (or to provide payments for others to purchase it), then farmers will benefit from the price floor, but taxpayers and consumers of food will pay the costs. Numerous proposals have been offered for reducing farm subsidies. In many countries, however, political support for subsidies for farmers remains strong. Either because this is viewed by the population as supporting the traditional rural way of life or because of the lobbying power of the agro-business industry.

For more detail on the effects price ceilings and floors have on demand and supply, see the following Application feature.

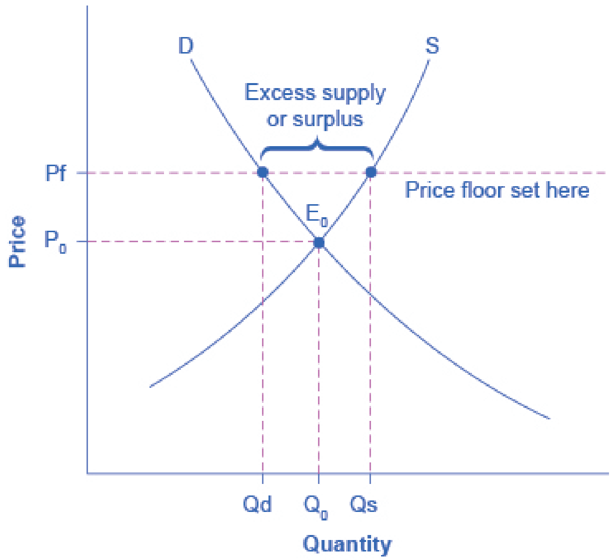


Figure 3.22. European Wheat Prices: A Price Floor Example. The intersection of demand (D) *and supply* (S) *would be at the equilibrium point E0.*

Do price ceilings and floors change demand or supply?

Neither price ceilings nor price floors cause demand or supply to change. They simply set a price that limits what can be legally charged in the market. Remember, changes in price do not cause demand or supply to change. Price ceilings and price floors can cause a different choice of quantity demanded along a demand curve, but they do not move the demand curve. Price controls can cause a different choice of quantity supplied along a supply curve, but they do not shift the supply curve.

DEMAND AND SUPPLY AS A SOCIAL ADJUSTMENT MECHANISM

The demand and supply model emphasizes that prices are not set only by demand or only by supply, but by the interaction between the two. In 1890, the famous economist **Alfred Marshall** wrote that asking whether supply or demand determined a price was like arguing "whether it is the upper or the under blade of a pair of scissors that cuts a piece of paper." The answer is that both blades of the demand and supply scissors are always involved.

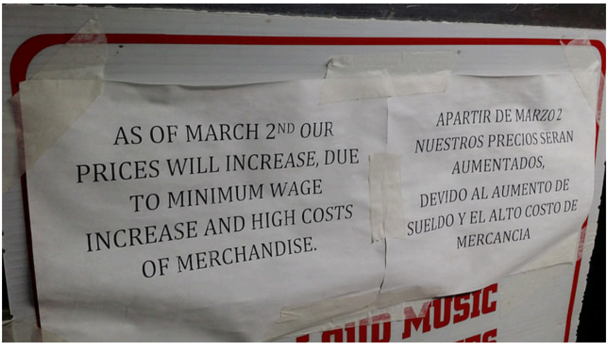


Image of store publicizing their price increases. By George Kelley. Accessed via Flickr and licensed under a CC-BY-ND license.

The adjustments of equilibrium price and quantity in a market-oriented economy often occur without much government direction or oversight. If the coffee crop in Brazil suffers a terrible frost, then the supply curve of coffee shifts to the left and the price of coffee rises. Some people—call them the coffee addicts—continue to drink coffee and pay the higher price. Others switch to tea or soft drinks. No government commission is needed to figure out how to adjust coffee prices, which companies will be allowed to process the remaining supply, which supermarkets in which cities will get how much coffee to sell, or which consumers will ultimately be allowed to drink the brew. Such adjustments in response to price changes happen all the time in a market economy, often so smoothly and rapidly that we barely notice them.

Think for a moment of all the seasonal foods that are available and inexpensive at certain times of the year, like fresh corn in midsummer, but more expensive at other times of the year. People alter their diets and restaurants alter their menus in response to these fluctuations in prices without fuss or fanfare. For both the U.S. economy and the world economy as a whole, markets—that is, demand and supply—are the primary social mechanism for answering the basic questions about what is produced, how it is produced, and for whom it is produced.

Self Check: Price Floors

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/723

104. Case in Point: Organic Foods

Why Can We Not Get Enough of Organic?

Organic food is grown without synthetic pesticides, chemical fertilizers or genetically modified seeds. In recent decades, the demand for organic products has increased dramatically. The Organic Trade Association reported sales increased from $1 billion in 1990 to $31.5 billion in 2011, more than 90% of which were sales of food products.

Why, then, are organic foods more expensive than their conventional counterparts? The answer is a clear application of the theories of supply and demand. As people have learned more about the harmful effects of chemical fertilizers, growth hormones, pesticides and the like from large-scale factory farming, our tastes and preferences for safer, organic foods have increased. This change in tastes has been reinforced by increases in income, which allow people to purchase pricier products, and has made organic foods more mainstream. This has led to an increased demand for organic foods. Graphically, the demand curve has shifted right, and we have moved up along the supply curve as producers have responded to the higher prices by supplying a greater quantity.

In addition to the movement along the supply curve, we have also had an increase in the number of farmers converting to organic farming over time. This is represented by a shift to the right of the supply curve. Since both demand and supply have shifted to the right, the resulting equilibrium quantity of organic foods is definitely higher, but the price will only fall when the increase in supply is larger than the increase in demand. We may need more time before we see lower prices in organic foods. Since the production costs of these foods may remain higher than conventional farming, because organic fertilizers and pest management techniques are more expensive, they may never fully catch up with the lower prices of non-organic foods.

As a final, specific example: The Environmental Working Group's "Dirty Dozen" list of fruits and vegetables, which test high for pesticide residue even after washing, was released in April 2013. The inclusion of strawberries on the list has led to an increase in demand for organic strawberries, resulting in both a higher equilibrium price and quantity of sales.

105. Self Check: Impact of Binding Price Ceilings or Price Floors

Check Your Understanding

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/627

106. Outcome: Tax Incidence

What you'll learn to do: explain how the price elasticities of demand and supply affect the incidence of a sales tax.

In this outcome, you'll learn about who bears the burden of a tax. Imagine a $1 tax on every barrel of apples an apple farmer produces. If the product (apples) is price inelastic to the consumer the farmer is able to pass the entire tax on to consumers of apples by raising the price by $1. In this example, consumers bear the entire burden of the tax; the tax incidence falls on consumers. On the other hand, if the apple farmer is unable to raise prices because the product is price elastic the farmer has to bear the burden of the tax or face decreased revenues: the tax incidence falls on the farmer. If the apple farmer can raise prices by an amount less than $1, then consumers and the farmer are sharing the tax burden.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Tax Incidence
- Self Check: Tax Incidence

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

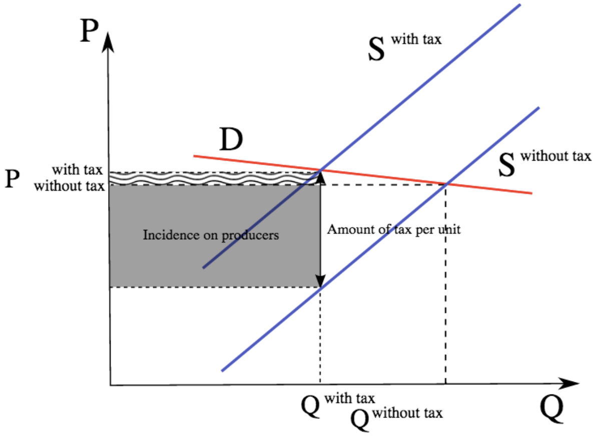
107. Reading: Tax Incidence

Tax Incidence

Tax incidence is the effect a particular tax has on the two parties of a transaction; the producer that makes the good and the consumer that buys it. The burden of the tax is not dependent on whether the state collects the revenue from the producer or consumer, but on the price elasticity of supply and the price elasticity of demand. To understand how elasticities influence tax incidence, it is important to consider the two extreme scenarios and how the tax burden is distributed between the two parties.

Inelastic supply, elastic demand

In this situation, because supply is inelastic, the firm will produce the same quantity no matter what the price. Because demand is elastic, the consumer is very sensitive to price. A small increase in price leads to a large drop in the quantity demanded. The imposition of the tax causes the market price to increase and the quantity demanded to decrease. Because consumption is elastic, the price consumers pay doesn't change very much. Because production is inelastic, the amount sold changes significantly. The producer is unable to pass the tax onto the consumer and the tax incidence falls on the producer.



Tax Incidence of Producer. *In a scenario with inelastic supply and elastic demand, the tax burden falls disproportionately on suppliers.*

Elastic supply, inelastic demand

Consumption is inelastic, so the consumer will consume near the same quantity no matter the price. The producer will be able to produce the same amount of the good, but will be able to increase the price by the amount of the tax. As a result, the entirety of the tax will be borne by the consumer.

Similarly elastic supply and demand

Generally consumers and producers are neither perfectly elastic or inelastic, so the tax burden is shared between the two parties in varying proportions. If one party is comparatively more inelastic than the other, they will pay the majority of the tax.

Increasing tax

If the government increases the tax on a good, that shifts the supply curve to the left, the consumer price increases, and sellers' price decreases. A tax increase does not affect the demand curve, nor does it make supply or demand more or less elastic. This potential increase in tax could be called marginal, because it is a tax in addition to existing levies.

Summary

- When supply is inelastic and demand is elastic, the tax incidence falls on the producer.
- When supply is elastic and demand is inelastic, the tax incidence falls on the consumer.
- Tax incidence is the analysis of the effect a particular tax has on the two parties of a transaction; the producer that makes the good and the consumer that buys it.
- A marginal tax is an increase in a tax on a good that shifts the supply curve to the left, increases the consumer price, and decreases the price for the sellers.

Self Check: Tax Incidence

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether

to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/628

108. Outcome: Taxation

What you'll learn to do: define progressive, proportional, and regressive taxes

"Our new Constitution is now established, and has an appearance that promises permanency; but in this world nothing can be said to be certain, except **death and taxes**."—Benjamin Franklin, IN A LETTER TO JEAN-BAPTISTE LEROY, 1789

In this section, you will take a closer look at the types of government taxes and examine why they are so certain in our lives.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Financing Government
- Reading: Types of Taxes
- Reading: Taxation
- Self Check: Taxation

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

109. Reading: Financing Government

If government services are to be provided, people must pay for them. The primary source of government revenue is taxes. In this section we examine the principles of taxation, compare alternative types of taxes, and consider the question of who actually bears the burden of taxes.

In addition to imposing taxes, governments obtain revenue by charging user fees, which are fees levied on consumers of government-provided services. The tuition and other fees charged by public universities and colleges are user fees, as are entrance fees at national parks. Finally, government agencies might obtain revenue by selling assets or by holding bonds on which they earn interest.

Principles of Taxation

Virtually anything can be taxed, but what should be taxed? Are there principles to guide us in choosing a system of taxes?

Jean-Baptiste Colbert, a minister of finance in seventeenth-century France, is generally credited with one of the most famous principles of taxation:

> *The art of taxation consists in so plucking the goose as to obtain the largest possible amount of feathers with the smallest possible amount of hissing.*

Economists, who do not typically deal with geese, cite two criteria for designing a tax system. The first is based on the ability of people to pay taxes and the second focuses on the benefits they receive from particular government services.

Ability to Pay

The ability-to-pay principle holds that people with more income should pay more taxes. As income rises, the doctrine asserts, people are able to pay more for public services; a tax system should therefore be constructed so that taxes rise too. Wealth, the total of assets less liabilities, is sometimes used as well as income as a measure of ability to pay.

The ability-to-pay doctrine lies at the heart of tax systems that link taxes paid to income received. The relationship between taxes and income may take one of three forms: taxes can be regressive, proportional, or progressive.

Regressive Tax

A regressive tax is one that takes a higher percentage of income as income falls. Taxes on cigarettes, for example, are regressive. Cigarettes are an inferior good—their consumption falls as incomes rise. Thus, people with lower incomes spend more on cigarettes than do people with higher incomes. The cigarette taxes paid by low-income people represent a larger share of their income than do the cigarette taxes paid by high-income people and are thus regressive.

Proportional Tax

A proportional tax is one that takes a fixed percentage of income. Total taxes rise as income rises, but taxes are equal to the same percentage no matter what the level of income. Some people argue that the U.S. income tax system should be changed into a flat tax system, a tax that would take the same percentage of income from all taxpayers. Such a tax would be a proportional tax.

Progressive Tax

A progressive tax is one that takes a higher percentage of income as income rises. The federal income tax is an example of a progressive tax. Table 15.1 shows federal income tax rates for various brackets of income for a family of four in 2007. Such a family paid no income tax at all if its income fell below $24,300. At higher income levels, families faced a higher percentage tax rate. Any income over $374,000, for example, was taxed at a rate of 35%. Whether or not to make the tax system more progressive was a major debating point during the U.S. presidential election of 2008.

***Table* 15.1 Federal Income Tax Brackets, 2007**

2007 adjusted gross income (family of four)	Personal income tax rate applied to bracket
Less than $24,300	Zero (family may receive earned income credit)
$24,300–$88,000	15%
$88,000–152,800	25%
$152,800–$220,150	28%
$220,150–$374,000	33%
Greater than $374,000	35%

The federal income tax is progressive. The percentage tax rate rises as adjusted gross income rises.

While a pure flat tax would be proportional, most proposals for such a tax would exempt some income from taxation. Suppose, for example, that households paid a "flat" tax of 20% on all income over $40,000 per year. This tax would be progressive. A household with an income of $25,000 per year would pay no tax. One with an income of $50,000 per year would pay a tax of $2,000 (.2 times $10,000), or 4% of its income. A household with an income of $100,000 per year would pay a tax of $12,000 (.2 times $60,000) per year, or 12% of its income. A flat tax with an income exemption would thus be a progressive tax.

Benefits Received

An alternative criterion for establishing a tax structure is the benefits-received principle, which holds that a tax should be based on the benefits received from the government services funded by the tax.

Local governments rely heavily on taxes on property, in large part because the benefits of many local services, including schools, streets, and the provision of drainage for wastewater, are reflected in higher property values. Suppose, for example, that public schools in a particular area are especially good. People are willing to pay more for houses served by those schools, so property values are higher; property owners benefit from better schools. The greater their benefit, the greater the property tax they pay. The property tax can thus be viewed as a tax on benefits received from some local services.

User fees for government services apply the benefits-received principle directly. A student paying tuition, a visitor paying an entrance fee at a national park, and a motorist paying a highway toll are all paying to consume a publicly provided service; they are thus paying directly for something from which they expect to benefit. Such fees can be used only for goods for which exclusion is possible; a user fee could not be applied to a service such as national defense.

Income taxes to finance public goods may satisfy both the ability-to-pay and benefits-received principles. The demand for public goods generally rises with income. Thus, people with higher incomes benefit more from public goods. The benefits-received principle thus suggests that taxes should rise with income, just as the ability-to-pay principle does. Consider, for example, an effort financed through income taxes by the federal government to clean up the environment. People with higher incomes will pay more for the cleanup than people with lower incomes, consistent with the ability-to-pay principle. Studies by economists consistently show that people with higher incomes have a greater demand for environmental improvement than do people with lower incomes—a clean environment is a normal good. Requiring people with higher incomes to pay more for the cleanup can thus be justified on the benefits-received principle as well.

Certainly taxes cannot respond precisely to benefits received. Neither the ability-to-pay nor the benefits-received doctrine gives us a recipe for determining just what each person "should" pay in taxes, but these doctrines give us a framework for thinking about the justification for particular taxes.

110. Reading: Types of Taxes

Types of Taxes

It is hard to imagine anything that has not been taxed at one time or another. Windows, closets, buttons, junk food, salt, death—all have been singled out for special taxes. In general, taxes fall into one of four primary categories.

Income taxes are imposed on the income earned by a person or firm; property taxes are imposed on assets; sales taxes are imposed on the value of goods sold; and excise taxes are imposed on specific goods or services. Figure 15.1 shows the major types of taxes financing all levels of government in the United States.

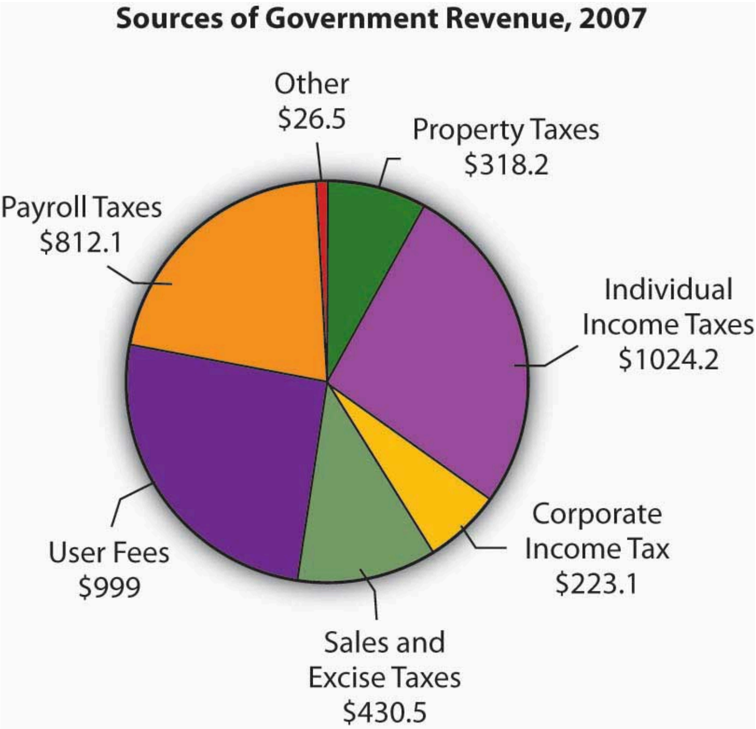


Figure 15.1 *Sources of Government Revenue, 2007. The chart shows sources of revenue for federal, state, and local governments in the United States. The data omit revenues from government-owned utilities and liquor stores. All figures are in billions of dollars. Source: U.S. Bureau of the Census, Statistical Abstract of US, 2008 (online) Tables 422 and 461.*

Personal Income Taxes

The federal personal income tax is the largest single source of tax revenue in the United States; most states and many cities tax income as well. All income tax systems apply a variety of exclusions to a taxpayer's total income before arriving at taxable income, the amount of income that is actually subject to the tax. In the U.S. federal income tax system, for example, a family deducted $3,200 from total income earned in 2005 for each member of the family as part of its computation of taxable income.

Income taxes can be structured to be regressive, proportional, or progressive. Income tax systems in use today are progressive.

In analyzing the impact of a progressive tax system on taxpayer choice, economists focus on the marginal tax rate. This is the tax rate that would apply to an additional $1 of taxable income earned. Suppose an individual was earning taxable income of $8,025 and paid federal income taxes of $802.50, or 10% of taxable income (we are ignoring exemptions that would eliminate taxes for such an individual). If the taxpayer were to receive $100 more of taxable income, however, that $100 would be taxed at a rate of 15%, the rate that applied in 2008 to taxable incomes between $8,025–$32,550 for individuals. That person thus faced a marginal tax rate of 15%.

Economists argue that choices are made at the margin; it is thus the marginal tax rate that is most likely to affect decisions. Say that the individual in our example is considering taking on additional work that would increase his or her income to $15,025 per year. With a marginal tax rate of 15%, the individual would keep $5,950 of the additional $7,000 earned. It is that $5,950 that the individual will weigh against the opportunity cost in forgone leisure in deciding whether to do the extra work.

Property Taxes

Property taxes are taxes imposed on assets. Local governments, for example, generally impose a property tax on business and personal property. A government official (typically a local assessor) determines the property's value, and a proportional tax rate is then applied to that value.

Property ownership tends to be concentrated among higher income groups; economists generally view property taxes as progressive. That conclusion, however, rests on assumptions about who actually pays the tax, an issue examined later in this module.

Sales Taxes

Sales taxes are taxes imposed as a percentage of firms' sales and are generally imposed on retail sales. Some items, such as food and medicine, are often exempted from sales taxation.

People with lower incomes generally devote a larger share of their incomes to consumption of goods covered by sales taxes than do people with higher incomes. Sales taxes are thus likely to be regressive.

Excise Taxes

An excise tax is imposed on specific items. In some cases, excise taxes are justified as a way of discouraging the consumption of demerit goods, such as cigarettes and alcoholic beverages. In other cases, an excise tax is a kind of benefits-received tax. Excise taxes on gasoline, for example, are typically earmarked for use in building and maintaining highways, so that those who pay the tax are the ones who benefit from the service provided.

The most important excise tax in the United States is the payroll tax imposed on workers' earnings. In 2007, the payroll tax was 12.4% and was levied on incomes up to $97,500. The Medicare portion of the payroll tax, 2.9%, was levied on all earned wages without limit. Half of the payroll tax is charged to employers, half to employees. The proceeds of this excise on payrolls finance Social Security and Medicare benefits. Almost two-thirds of U. S. households pay more in payroll taxes than in any other taxes.

Tax Incidence Analysis

Next time you purchase an item at a store, notice the sales tax imposed by your state, county, and city. The clerk rings up the total, then adds up the tax. The store is the entity that "pays" the sales tax, in the sense that it sends the money to the government agencies that imposed it, but you are the one who actually foots the bill—or are you? Is it possible that the sales tax affects the price of the item itself?

These questions relate to tax incidence analysis, a type of economic analysis that seeks to determine where the actual burden of a tax rests. Does the burden fall on consumers, workers, owners of capital, owners of natural resources, or owners of other assets in the economy? When a tax imposed on a good or service increases the price by the amount of the tax, the burden of the tax falls on consumers. If instead it lowers wages or lowers prices for some of the other factors of production used in the production of the good or service taxed, the burden of the tax falls on owners of these factors. If the tax does not change the product's price or factor prices, the burden falls on the owner of the firm—the owner of capital. If prices adjust by a fraction of the tax, the burden is shared.

Figure 15.2 gives an example of tax incidence analysis. Suppose *D1* and *S1* are the demand and supply curves for beef. The equilibrium price is $3 per pound; the equilibrium quantity is 3 million pounds of beef per day. Now suppose an excise tax of $2 per pound of beef is imposed. It does not matter whether the tax is levied on buyers or on sellers of beef; the important thing to see is that the tax drives a $2 per pound "wedge" between the price buyers pay and the price sellers receive. This tax is shown as the vertical green line in the exhibit; its height is $2.

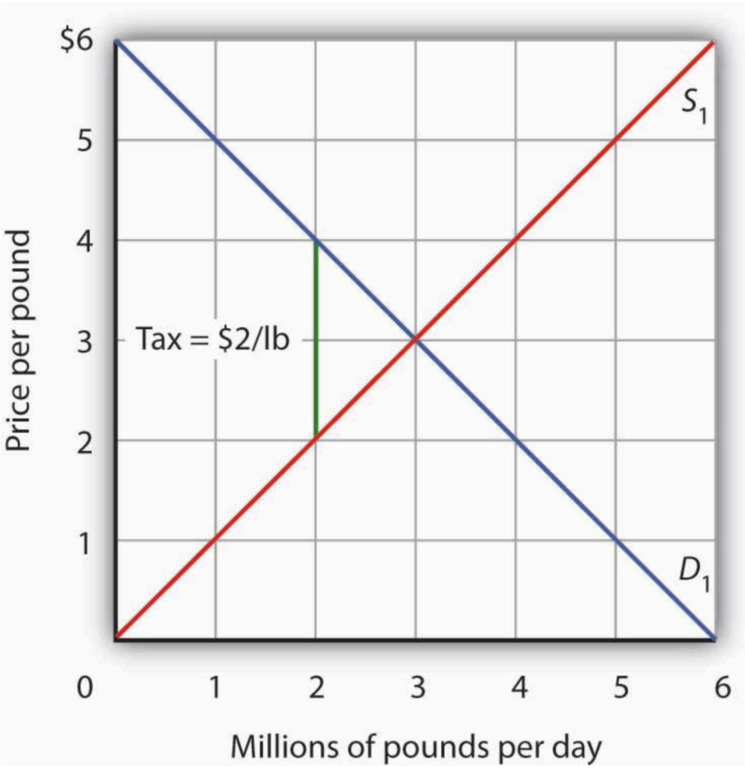


Figure 15.2 *Tax Incidence in the Model of Demand and Supply. Suppose the market price of beef is $3 per pound; the equilibrium quantity is 3 million pounds per day. Now suppose an excise tax of $2 per pound is imposed, shown by the vertical green line. We insert this tax wedge between the demand and supply curves. It raises the market price to $4 per pound, suggesting that buyers pay half the tax in the form of a higher price. Sellers receive a price of $2 per pound; they pay half the tax by receiving a lower price. The equilibrium quantity falls to 2 million pounds per day.*

We insert our tax "wedge" between the demand and supply curves. In our example, the price paid by buyers rises to $4 per pound. The price received by sellers falls to $2 per pound; the other $2 goes to the government. The quantity of beef demanded and supplied falls to 2 million pounds per day. In this case, we conclude that buyers bear half the burden of the tax (the price they pay rises by $1 per pound), and sellers bear the other half (the price they receive falls by $1 per pound). In addition to the change in price, a further burden of the tax results from the reduction in consumer and in producer surplus. We have not shown this reduction in the graph.

Figure 15.3 shows how tax incidence varies with the relative elasticities of demand and supply. All four panels show markets with the same initial price, P1, determined by the intersection of demand D1and supply S1. We impose an excise tax, given by the vertical green line. As before, we insert this tax wedge between the demand and supply curves. We assume the amount of the tax per unit is the same in each of the four markets.

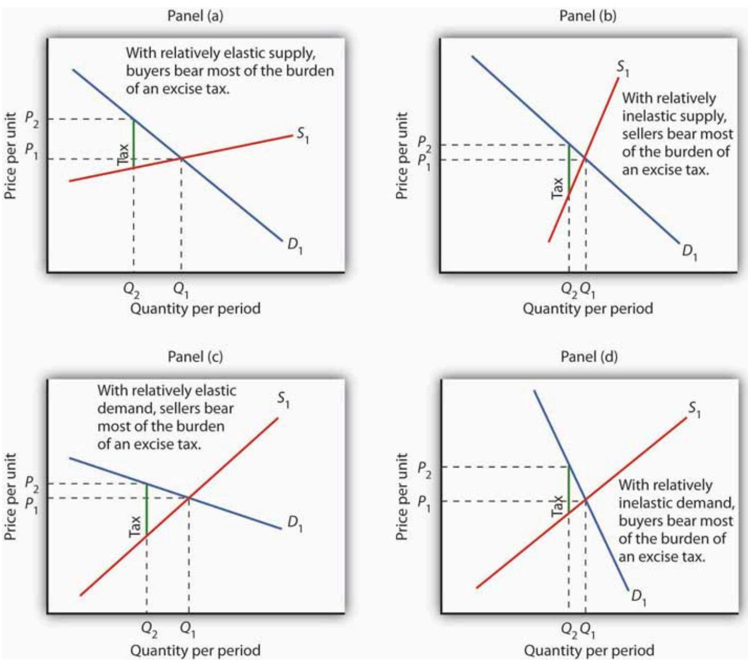


***Figure 15.3** Tax Incidence and the Elasticity of Demand and of Supply. We show the effect of an excise tax, given by the vertical green line, in the same way that we did in Figure 15.2. We see that buyers bear most of the burden of such a tax in cases of relatively elastic supply (Panel (a)) and of relatively inelastic demand (Panel (d)). Sellers bear most of the burden in cases of relatively inelastic supply (Panel (b)) and of relatively elastic demand (Panel (c)).*

In Panel (a), we have a market with a relatively elastic supply curve S1. When we insert our tax wedge, the price rises to P2; the price increase is nearly as great as the amount of the tax. In Panel (b), we have the same demand curve as in Panel (a), but with a relatively inelastic supply curve S2. This time the price paid by buyers barely rises; sellers bear most of the burden of the tax. When the supply curve is relatively elastic, the bulk of the tax burden is borne by buyers. When supply is relatively inelastic, the bulk of the burden is borne by sellers.

Panels (c) and (d) of the exhibit show the same tax imposed in markets with identical supply curves S1. With a relatively elastic demand curve D1in Panel (c) (notice that we are in the upper half, that is, the elastic portion of the curve), most of the tax burden is borne by sellers. With a relatively inelastic demand curve D1in Panel (d) (notice that we are in the lower half, that is, the inelastic portion of the curve), most of the burden is borne by buyers. If demand is relatively elastic, then sellers bear more of the burden of the tax. If demand is relatively inelastic, then buyers bear more of the burden.

The Congressional Budget Office (CBO) has prepared detailed studies of the federal tax system. Using the tax laws in effect in August 2004, it ranked the U.S. population according to income and then divided the population into quintiles (groups containing 20% of the population). Then, given the federal tax burden imposed by individual income taxes, payroll taxes for social insurance, corporate income taxes, and excise taxes on each quintile and the income earned by people in that quintile, it projected the average tax rate facing that group in 2006. The study assigned taxes on the basis of who bears the burden, not on who pays the tax. For example, many studies argue that, even though businesses pay half of the payroll taxes, the burden of payroll taxes actually falls on households. The reason is that the supply curve of labor is relatively inelastic, as shown in Panel (b) of Figure 15.3. Taking these adjustments into account, the CBO's results, showing progressivity in federal taxes, are reported in Table 15.2.

Table 15.2 Federal Tax Burdens in the United States

Income category	Households (number, millions)	Average pretax comprehensive household income	Effective federal tax rate, 2006 (percent)
Lowest quintile	24.0	$18,568	5.6
Second quintile	22.8	$42,619	12.1
Middle quintile	23.3	$64,178	15.7
Fourth quintile	23.2	$94,211	19.8
Highest quintile	24.3	$227,677	26.5
All quintiles	118.3	$89,476	21.6

In a regressive tax system, people in the lowest quintiles face the highest tax rates. A proportional system imposes the same rates on everyone; a progressive system imposes higher rates on people in higher deciles. The table gives estimates by the CBO of the burden on each quintile of federal taxes in 2006. As you can see, the tax structure in the United States is progressive.

KEY TAKEAWAYS

1. The primary principles of taxation are the ability-to-pay and benefits-received principles.
2. The percentage of income taken by a regressive tax rises as income falls. A proportional tax takes a constant percentage of income regardless of income level. A progressive tax takes a higher percentage of income as taxes as incomes rise.
3. The marginal tax rate is the tax rate that applies to an additional dollar of income earned.
4. Tax incidence analysis seeks to determine who ultimately bears the burden of a tax.

5. The major types of taxes are income taxes, sales taxes, property taxes, and excise taxes.
6. Buyers bear most of the burden of an excise tax when supply is relatively elastic and when demand is relatively inelastic; sellers bear most of the burden when supply is relatively inelastic and when demand is relatively elastic.
7. The federal tax system in the United States is progressive.

111. Worked Example: Price Controls

Price Controls

People often expect the government to solve problems that they seem unable to solve on their own. Sometimes this is effective and sometimes it is not. Price controls, either price ceilings or price floors, often have unanticipated side effects. Think about it—passing a law doesn't by itself make economic problems go away!

Such is the case with claims of price gouging, the charging of "excessively high" prices, which was exemplified by what occurred in the wake of Hurricane Katrina. Imposing a price ceiling below the equilibrium price may create as many problems as it solves. The basic problem is that the demand for bottled water is dramatically higher, since fresh water supplies were compromised, while the supply of bottled water was less as a result of storm damage.

The question is how to deal with the shortage, that is, how to allocate the limited supply of bottled water among competing needs and wants. Figure 1 depicts the example of price control for water after Hurricane Katrina.

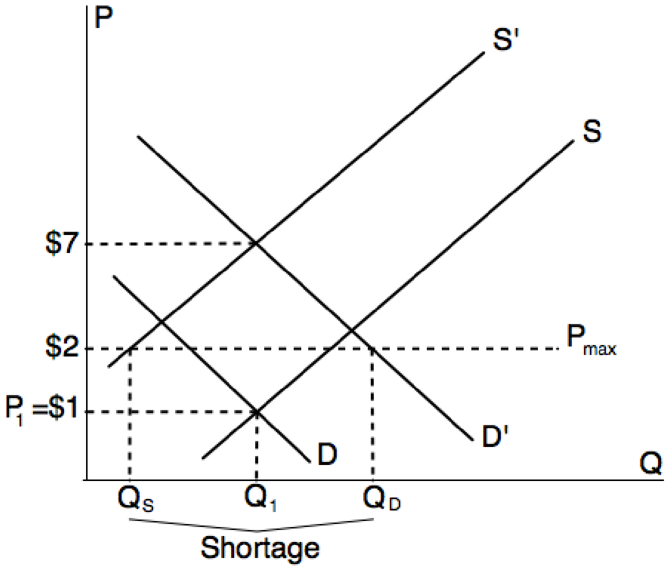


Figure 1. *Price control for water after Hurricane Katrina.*

When a price ceiling reduces the legal price of a product, businesses have less incentive to supply the product. Economically speaking, the law of supply says that at lower prices, the quantity supplied will be lower. At the same time, the law of demand states that at a lower price, the quantity demanded will be higher. This can be seen clearly in the graph. So who gets the limited supply? As Shakespeare said, that is the question.

Unfortunately, there is no clear answer to this. It could be first come, first serve. It could be friends of the seller. In many cases, what results are under-the-table payments by consumers willing to violate the law. What is certain is that less bottled water gets to consumers than would be the case if the price were allowed to rise. Many would argue that this shortfall is not the best outcome.

112. Reading: Taxation

Taxation

There are two main categories of taxes: those collected by the federal government and those collected by state and local governments. What percentage is collected and what that revenue is used for varies greatly. The following sections will briefly explain the taxation system in the United States.

Federal Taxes

Just as many Americans erroneously think that federal spending has grown considerably, many also believe that taxes have increased substantially. The top line of Figure 16.5 shows total federal taxes as a share of GDP since 1960. Although the line rises and falls, it typically remains within the range of 17% to 20% of GDP, except for 2009, when taxes fell substantially below this level, due to recession.

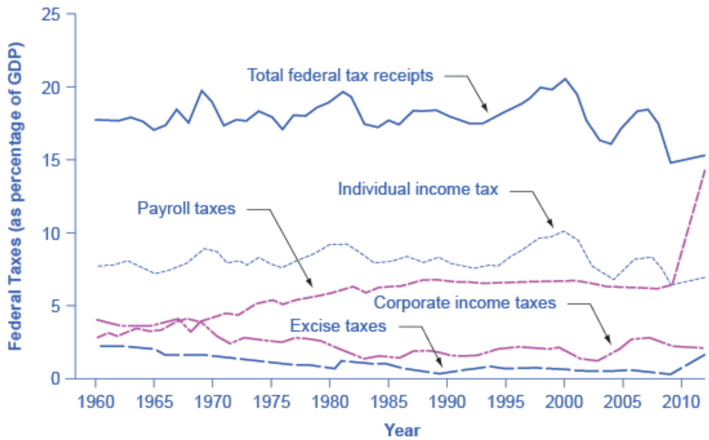


Figure 16.5. *Federal Taxes, 1960–2012. Federal tax revenues have been about 17–20% of GDP during most periods in recent decades. The primary sources of federal taxes are individual income taxes and the payroll taxes that finance Social Security and Medicare. Corporate income taxes, excise taxes, and other taxes provide smaller shares of revenue. (Source: Economic Report of the President, Tables B-81 and B-1, http://www.gpo.gov/fdsys/pkg/ERP-2013/content-deta…)*

Figure 16.5 also shows the patterns of taxation for the main categories of taxes levied by the federal government: personal income taxes, payroll taxes, corporate income taxes, and excise taxes. When most people think of taxes levied by the federal government, the first tax that comes to mind is the *individual income tax* that is due every year on April 15 (or the first business day after). The personal income tax is the largest single source of federal government revenue, but it still represents less than half of federal tax revenue.

The second largest source of federal revenue is the *payroll tax*, which provides funds for Social Security and Medicare. Payroll taxes have increased steadily over time. Together, the personal income tax and the payroll tax accounted for about 84% of federal tax revenues in 2012. Although personal income tax revenues account for more total revenue than the payroll tax, nearly three-quarters of households pay more in payroll taxes than in income taxes.

The income tax is a *progressive tax*, which means that the tax rates increase as a household's income increases. Taxes also vary with marital status, family size, and other factors. The *marginal tax rates* (the tax that must be paid on all yearly income) for a single taxpayer range from 10% to 35%, depending on income, as explained below.

HOW DOES THE MARGINAL RATE WORK?

Suppose that a single taxpayer's income is $35,000 per year. Also suppose that income from $0 to $9,075 is taxed at 10%, income from $9,075 to $36,900 is taxed at 15%, and, finally, income from $36,900 and beyond is taxed at 25%. Since this person earns $35,000, their marginal tax rate is 15%.

The key fact here is that the federal income tax is designed so that tax rates increase as income increases, up to a certain level. The payroll taxes that support Social Security and Medicare are designed in a different way. First, the payroll taxes for Social Security are imposed at a rate of 12.4% up to a certain wage limit, set at $117,900 in 2014. Medicare, on the other hand, pays for elderly healthcare, and is fixed at 2.9%, with no upper ceiling.

In both cases, the employer and the employee split the payroll taxes. An employee only sees 6.2% deducted from his paycheck for Social Security, and 1.45% from Medicare. However, as economists are quick to point out, the employer's half of the taxes are probably passed along to the employees in the form of lower wages, so in reality, the worker pays all of the payroll taxes.

The Medicare payroll tax is also called a *proportional tax*; that is, a flat percentage of all wages earned. The Social Security payroll tax is proportional up to the wage limit, but above that level it becomes a *regressive tax*, meaning that people with higher incomes pay a smaller share of their income in tax.

The third-largest source of federal tax revenue, as shown in Figure 16.5 is the *corporate income tax*. The common name for corporate income is "profits." Over time, corporate income tax receipts have declined as a share of GDP, from about 4% in the 1960s to an average of 1% to 2% of GDP in the first decade of the 2000s.

The federal government has a few other, smaller sources of revenue. It imposes an *excise tax*—that is, a tax on a particular good—on gasoline, tobacco, and alcohol. As a share of GDP, the amount collected by these taxes has stayed nearly constant over time, from about 2% of GDP in the 1960s to roughly 3% by 2012, according to the nonpartisan Congressional Budget Office. The government also imposes an *estate and gift tax* on people who pass large amounts of assets to the next generation—either after death or during life in the form of gifts. These estate and gift taxes collected about 0.2% of GDP in the first decade of the 2000s. By a quirk of legislation, the estate and gift tax was repealed in 2010, but reinstated in 2011. Other federal taxes, which are also relatively small in magnitude, include tariffs collected on imported goods and charges for inspections of goods entering the country.

State and Local Taxes

At the state and local level, taxes have been rising as a share of GDP over the last few decades to match the gradual rise in spending, as Figure 16.6 illustrates. The main revenue sources for state and local governments are sales taxes, property taxes, and revenue passed along from the federal government, but many state and local governments also levy personal and corporate income taxes, as well as impose a wide variety of fees and charges. The specific sources of tax revenue vary widely across state and local governments. Some states rely more on property taxes, some on sales taxes, some on income taxes, and some more on revenues from the federal government.

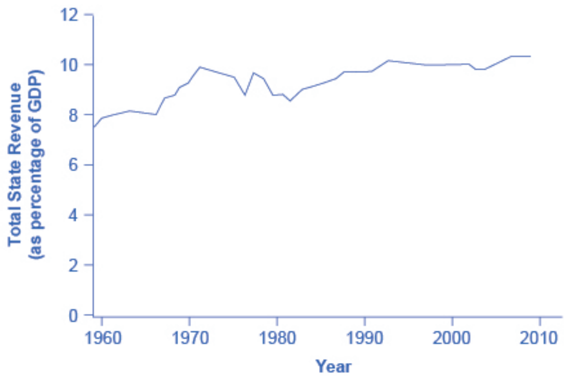


Figure 16.6. *State and Local Tax Revenue as a Share of GDP, 1960–2010. State and local tax revenues have increased to match the rise in state and local spending. (Source: Economic Report of the President, Tables B-85 and B-1, http://www.gpo.gov/fdsys/pkg/ERP-2013/content-deta…)*

Self Check: Taxation

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the three Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/629

113. Putting It Together: Government Action

Summary

This module was an application of the theories of supply and demand. You will notice that once you learn a theory in this course, you will use it over and over again, so the theories are important to learn (click here for a review of theories from the earlier module). The goal of this module was to explore the impacts of government intervention in markets. You learned how to:

- Define and analyze the impacts of price floors like the minimum wage
- Define and analyze the impacts of price ceilings like rent control
- Describe the incidence of various transaction-based taxes, that is, who bears the burden of the tax?
- Identify the types of taxes and the sources of federal income
- Define and analyze the impact of taxes as a percentage of one's income. In other words, is the burden on the rich, the poor or everyone equally?

Examples

Let's return to the example of the minimum wage. Careful analysis shows that imposition of, or increases in the minimum wage have significant distributional effects. In other words, there are winners and losers from the policy. The winners are workers who continue to have a job, but are now paid a higher salary. The losers are businesses who have to pay more for their employees. This increase in production costs will be passed on, in part to consumers who will end up paying higher prices for the businesses' products. So consumers lose also.

The big losers, though, are the people who had jobs at the lower wage, but lose them when the minimum wage is increased. Which employees are most likely to lose their jobs, the most experienced and skilled, or the least experienced and skilled? Don't forget that ultimately what matters is the size of these effects, which we learned in the last module are based on wage elasticities of supply and demand for labor and the price elasticity of demand for the firm's products. These are the technical details that policy analysts will look at before making any recommendations to decision makers.

114. Glossary: Government Action

price ceiling
a legal maximum price

price control
government laws to regulate prices instead of letting market forces determine prices

price floor
a legal minimum price

total surplus
see social surplus

115. Discussion: Junk Food and Government Action

Recently there has been discussion in the news about taxing junk food (soft drinks, for example) in an effort to reduce the incidence of obesity in the U.S. Do you think the demand for junk food is elastic or inelastic with respect to price.

Based on your knowledge of the price elasticity of demand, do you think the deadweight loss of a soda/junk-food tax would be relatively large or relatively small? Why? Do you think taxing junk food would be a good idea? Based on your analysis, would it really help reduce the number of obese people in the United States? Explain.

PART VII

MODULE: SURPLUS

116. Why It Matters: Surplus

Why use the concept of producer, consumer surplus, and total surplus to explain the outcomes of markets for individuals, firms, and society?

Students often see this topic on surplus as technical, but it's really fundamental to understanding economics if you realize what it's about. Economists believe that voluntary transactions (purchases and sales) are mutually beneficial. It is not the case that one side gains and the other side loses. Rather, transactions are positive sum games in which both parties are better off as a result. In principle, we can measure the gains to both parties. This measurement is the rationale for two important concepts: consumer surplus and producer surplus, which together make up economic (or social) surplus—the gain to society from the transaction. This is the subject of this module.

Surplus, in the context of this module, just means how good a deal a consumer got on a purchase, or how good a deal a producer got on a sale. That's it in a nutshell.

Sale Image by Tim Parkinson, CC-BY.

For consumers, this is often highlighted by a sale when deals become bigger. Sales promotions bring in customers who wouldn't pay the normal price. But they also allow customers who would have purchased anyway to get an even better deal. Watch this BBC video about Cyber Monday to find out more about how consumers react to large sales.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=140#oembed-1

How big a deal do consumers get together from Cyber Monday? We can answer this question by computing the consumer surplus.

What about businesses? Why do they run sales like Cyber Monday, and how much do they gain from them? The answer can be found by computing the producer surplus.

Let's see how it's done.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=140

LEARNING OUTCOMES

- Define and calculate consumer surplus
- Define and calculate producer surplus
- Define and calculate total surplus
- Use the concepts of consumer, producer and total surplus to explain why markets typically lead to efficient outcomes

117. Outcome: Consumer, Producer, and Total Surplus

What you'll learn to do: define and calculate consumer, producer, and total surplus; graphically illustrate consumer, producer, and total surplus

Imagine that you want to buy a new smartphone. You have been saving up money and are willing to pay $500 for you new phone. To your surprise, you go to the phone store and learn that recent changes in phone company legislation have lowered the price of your favorite phone to only $250. This is an obvious win for you, but from an economic standpoint, there is a surplus in the market. Consumers are willing to pay more than $250 for cell phones, so the equilibrium price is actually less than what consumers would be willing to pay. In this outcome, we will understand what happens to a market when there is a consumer, producer, or total surplus.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Surplus
- Self Check: Consumer, Producer, and Total Surplus

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

118. Reading: Surplus

Demand, Supply, and Efficiency

The familiar demand and supply diagram holds within it the concept of economic efficiency. One typical way that economists define efficiency is when it is impossible to improve the situation of one party without imposing a cost on another. Conversely, if a situation is inefficient, it becomes possible to benefit at least one party without imposing costs on others.

Efficiency in the demand and supply model has the same basic meaning: The economy is getting as much benefit as possible from its scarce resources and all the possible gains from trade have been achieved. In other words, the optimal amount of each good and service is being produced and consumed.

Consumer Surplus, Producer Surplus, Social Surplus

Consider a market for tablet computers, as shown in Figure 1. The equilibrium price is $80 and the equilibrium quantity is 28 million. To see the benefits to consumers, look at the segment of the demand curve above the equilibrium point and to the left. This portion of the demand curve shows that at least some demanders would have been willing to pay more than $80 for a tablet.

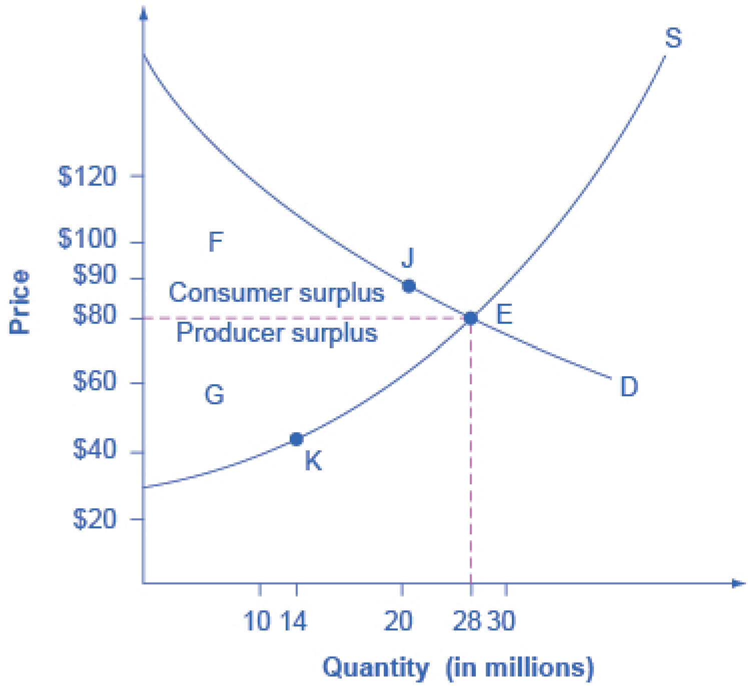


Figure 1: Consumer and Producer Surplus

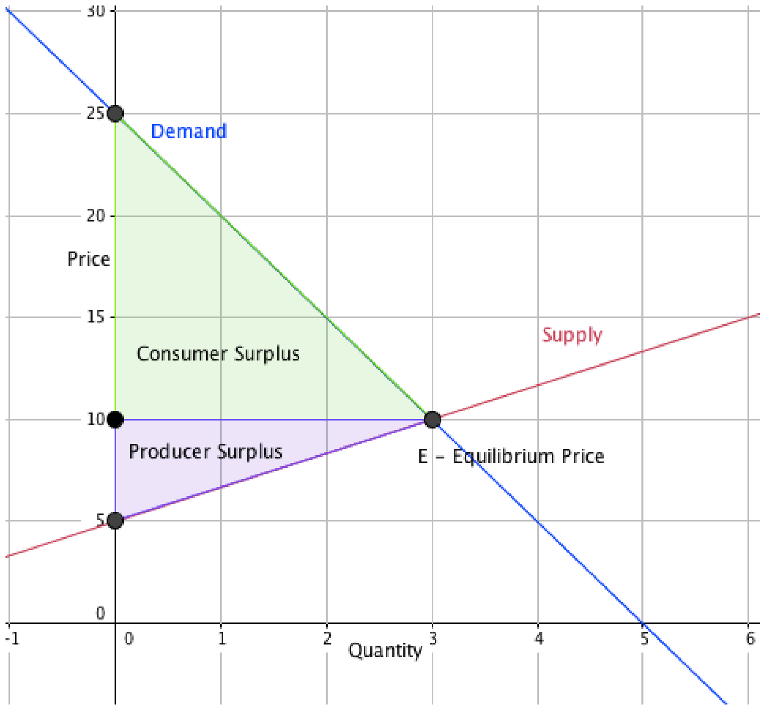
For example, point J shows that if the price was \$90, 20 million tablets would be sold. Those consumers who would have been willing to pay \$90 for a tablet based on the utility they expect to receive from it, but who were able to pay the equilibrium price of \$80, clearly received a benefit beyond what they had to pay for. Remember, the demand curve traces consumers' willingness to pay for different quantities. The amount that individuals would have been willing to pay, minus the amount that they actually paid, is called consumer surplus. Consumer surplus is the area labeled F—that is, the area above the market price and below the demand curve.

The somewhat triangular area labeled by F in the graph above shows the area of consumer surplus, which shows that the equilibrium price in the market was less than what many of the consumers were willing to pay. Point J on the demand curve shows that, even at the price of $90, consumers would have been willing to purchase a quantity of 20 million. The somewhat triangular area labeled by G shows the area of producer surplus, which shows that the equilibrium price received in the market was more than what many of the producers were willing to accept for their products. For example, point K on the supply curve shows that at a price of $45, firms would have been willing to supply a quantity of 14 million.

The supply curve shows the quantity that firms are willing to supply at each price. For example, point K in Figure 1 illustrates that, at $45, firms would still have been willing to supply a quantity of 14 million. Those producers who would have been willing to supply the tablets at $45, but who were instead able to charge the equilibrium price of $80, clearly received an extra benefit beyond what they required to supply the product. The amount that a seller is paid for a good minus the seller's actual cost is called producer surplus. In Figure 1, producer surplus is the area labeled G—that is, the area between the market price and the segment of the supply curve below the equilibrium.

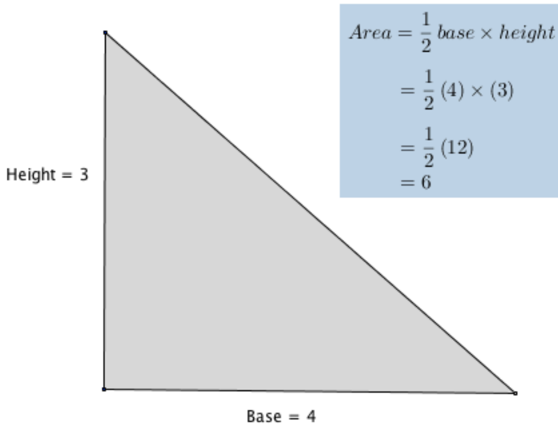
LINK IT UP

Example: Calculate consumer surplus given linear supply and demand curves.



In the sample market shown in the graph, equilibrium price is $10 and equilibrium quantity is 3 units. The consumer surplus area is highlighted above the equilibrium price line. This area can be calculated as the area of a triangle.

Recall that to find the area of a triangle, you will need to know it's base and height. Refer to the following example if you need a refresher.



Let's apply the calculation for the area of a triangle to our example market to see the added value that consumers will get for this item at the equilibrium price in our sample market.

Step 1 Define the base and height of the consumer surplus triangle.

The base of the consumer surplus triangle is 3 units long. Be careful when you define the height of this triangle, it is tempting to say it is 25, can you see why it isn't? The height is determined by the distance from the equilibrium price line and where the demand curve intersects the vertical axis. The height of the triangle begins at $10 and ends at $25, so it will be $25 – $10 = $15

$$b = 3$$
$$h = 15$$

Step 2 Apply the values for base and height to the formula for the area of a triangle.

A = $\frac{1}{2}b \times h$

A = $\frac{1}{2} 3 \times 15$

A = $\frac{1}{2} 45$

A = $\frac{45}{2} = 22.5$

By calculating the consumer surplus value, we can gain insight into the price elasticity of supply and demand. When demand is inelastic, consumer surplus is greater. Can you explain why? How would consumer surplus change as market price changes?

The sum of consumer surplus and producer surplus is social surplus, also referred to as economic surplus or total surplus. In Figure 1, social surplus would be shown as the area F + G. Social surplus is larger at equilibrium quantity and price than it would be at any other quantity. This demonstrates the economic efficiency of the market equilibrium. In addition, at the efficient level of output, it is impossible to produce greater consumer surplus without reducing producer surplus, and it is impossible to produce greater producer surplus without reducing consumer surplus.

Self Check: Consumer, Producer, and Total Surplus

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/724

119. Outcome: Surplus and Inefficiency

What you'll learn to do: use the concepts of consumer, producer, and total surplus to explain why markets typically lead to efficient outcomes

In this section, you'll build on your understanding of surplus from the previous outcome to examine the connection between types of surplus and the impact they have on the economy.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Inefficiency of Price Floors and Price Ceilings
- Self Check: Surplus and Inefficiency

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

120. Reading: Inefficiency of Price Floors and Price Ceilings

Inefficiency of Price Floors and Price Ceilings

The imposition of a price floor or a price ceiling will prevent a market from adjusting to its equilibrium price and quantity, and thus will create an inefficient outcome. But there is an additional twist here. Along with creating inefficiency, price floors and ceilings will also transfer some consumer surplus to producers, or some producer surplus to consumers.

Imagine that several firms develop a promising but expensive new drug for treating back pain. If this therapy is left to the market, the equilibrium price will be $600 per month and 20,000 people will use the drug, as shown in Figure 2a. The original level of consumer surplus is T + U and producer surplus is V + W + X. However, the government decides to impose a price ceiling of $400 to make the drug more affordable. At this price ceiling, firms in the market now produce only 15,000.

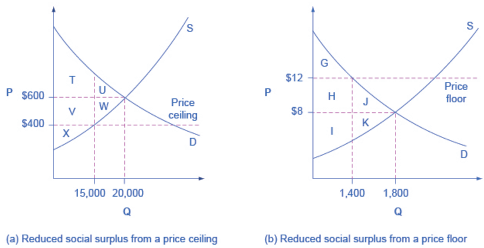


Figure 2. *(a) The original equilibrium price is $600 with a quantity of 20,000. Consumer surplus is T + U, and producer surplus is V + W + X. A price ceiling is imposed at $400, so firms in the market now produce only a quantity of 15,000. As a result, the new consumer surplus is T + V, while the new producer surplus is X. (b) The original equilibrium is $8 at a quantity of 1,800. Consumer surplus is G + H + J, and producer surplus is I + K. A price floor is imposed at $12, which means that quantity demanded falls to 1,400. As a result, the new consumer*

surplus is G, and the new producer surplus is H + I.

As a result, two changes occur. First, an inefficient outcome occurs and the total surplus of society is reduced. The loss in social surplus that occurs when the economy produces at an inefficient quantity is called deadweight loss. In a very real sense, it is like money thrown away that benefits no one. In Figure 2.a, the deadweight loss is the area U + W. When deadweight loss exists, it is possible for both consumer and producer surplus to be higher, in this case because the price control is blocking some suppliers and demanders from transactions they would both be willing to make.

A second change from the price ceiling is that some of the producer surplus is transferred to consumers. After the price ceiling is imposed, the new consumer surplus is T + V, while the new producer surplus is X. In other words, the price ceiling transfers the area of surplus (V) from producers to consumers. Note that the gain to consumers is less than the loss to producers, which is just another way of seeing the deadweight loss.

Efficiency and Price Floors and Ceilings

Figure 2.b shows a price floor example using a string of struggling movie theaters, all in the same city. The current equilibrium is $8 per movie ticket, with 1,800 people attending movies. The original consumer surplus is G + H + J, and producer surplus is I + K. The city government is worried that movie theaters will go out of business, reducing the entertainment options available to citizens, so it decides to impose a price floor of $12 per ticket. As a result, the quantity demanded of movie tickets falls to 1,400. The new consumer surplus is G, and the new producer surplus is H + I. In effect, the price floor causes the area H to be transferred from consumer to producer surplus, but also causes a deadweight loss of J + K.

This analysis shows that a price ceiling, like a law establishing rent controls, will transfer some producer surplus to consumers—which helps to explain why consumers often favor them. Conversely, a price floor like a guarantee that farmers will receive a certain price for their crops will transfer some consumer surplus to producers, which explains why producers often favor them. However, both price floors and price ceilings block some transactions that buyers and sellers would have been willing to make, and creates deadweight loss. Removing such barriers, so that prices and quantities can adjust to their equilibrium level, will increase the economy's social surplus.

Summary

Consumer surplus is the gap between the price that consumers are willing to pay, based on their preferences, and the market equilibrium price. Producer surplus is the gap between the price for which producers are willing to sell a product, based on their costs, and the market equilibrium price. Social surplus is the sum of consumer surplus and producer surplus. Total surplus is larger at the equilibrium quantity and price than it will be at any other quantity and price. Deadweight loss is loss in total surplus that occurs when the economy produces at an inefficient quantity.

Self Check: Surplus and Inefficiency

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/633

121. Glossary: Surplus

consumer surplus
: the extra benefit consumers receive from buying a good or service, measured by what the individuals would have been willing to pay minus the amount that they actually paid

deadweight loss
: the loss in social surplus that occurs when a market produces an inefficient quantity

economic surplus
: see social surplus

producer surplus
: the extra benefit producers receive from selling a good or service, measured by the price the producer actually received minus the price the producer would have been willing to accept

social surplus
: the sum of consumer surplus and producer surplus

122. Putting It Together: Surplus

Summary

The goal of this module was to use the concepts of consumer surplus, producer surplus, and total economic surplus to explain the outcomes of markets for individuals, firms and society. You learned how to:

- Define and calculate consumer surplus
- Define and calculate producer surplus
- Understand total economic surplus as the sum of consumer and producer surplus
- Use the concepts of consumer, producer and total surplus to explain why markets typically lead to efficient outcomes

Examples

Consider Groupon, a website which offers significant discounts on purchases at businesses people frequently use. It's not unusual to obtain 50% off the normal price. Why do customers like Groupon? Because it increases the consumer surplus they obtain on purchases.

Groupon Gift Card Image by Mike Mozart, CC-BY.

Why do businesses offer Groupon campaigns? Part of it is advertising, to attract customers who aren't familiar with those businesses. Some businesses offer regular Groupon deals. They must be doing this to increase their producer surplus (i.e., profit). This is likely part of a larger strategy, called *price discrimination*, which we will learn more about when we study the theory of the firm. For now, it is enough to understand that Groupon campaigns enhance producer surplus.

Since both consumer surplus and producer surplus increase, we can say that total economic (or social) surplus has increased. This is just another way of saying that transactions benefit both parties, or as economists would say, this is a more efficient outcome for society. Computing the additional consumer and producer surplus tells us by how much economic surplus has increased.

123. Discussion: Efficiency of Free Markets

Economists have long known that free markets are perfectly efficient. Consider Q1 in the graph, above. How much do consumers value one more unit of output (Q1 + 1)? How much do producers value one more unit of output? Use the concepts of consumer surplus and producer surplus to explain why Q2 is more efficient than Q1. Now use the concept of total economic surplus to explain why the equilibrium quantity is the most efficient. Why would Q4 be less efficient than the equilibrium quantity?

PART VIII

MODULE: UTILITY

124. Why It Matters: Utility

Why explain how consumer behavior shapes the demand curve with respect to utility and loss?

Consumption Collage. All images by Chris Potter, [CC-BY](). [Best Buy](), [Costco](), [Macy's](), [Home Depot]()

As a student you may be interested in maximizing your grade point average (GPA) in order to qualify for scholarships, transfer to a different college or university, or even go to graduate school. In order to make your GPA as high as you can, which of the following two strategies should you choose?

- Spend equal amounts of time studying for each of your classes, or
- Study more for some classes and less for others

Generally, the answer is that you should be studying more for some classes and less for others. Why?

This module introduces the "biggest bang for the buck" principle, which is one of the key principles of microeconomics. This principle explains how to make the best choices you can in conditions of scarcity. For example, if you only have a few hours you can spend studying each day, how should you divide them up between your different classes?

This module will also help you answer questions like:

- Why do people purchase more of something when its price falls?
- Why do people buy more goods and services when their budget increases?

The Consumer's Mind by Christopher Dombres, CC-BY.

In the first portion of this module, we will assume that consumers are economically rational – this is a key assumption of mainstream microeconomics. Toward the end of the module we relax this assumption as we consider an alternative approach called behavioral economics. Behavioral economics acknowledges that sometimes we don't make the best choices.

LEARNING OUTCOMES

- Define the concept of utility and satisfaction
- Explain how consumers maximize total utility within a given

income using the Utility Maximizing Rule
- Explain how consumer's utility changes when income or prices change
- Describe the behavioral economics approach to understanding decision making

125. Outcome: Defining Utility

What you'll learn to do: define the concept of utility and satisfaction.

In this outcome, you will learn to gauge the measure of satisfaction that a person gets from buying a product.

The specific things you'll learn to do in this section include:

- Differentiate between marginal utility and total utility

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Consumer Choices
- Reading: Consumption Choices
- Reading: Choosing with Marginal Utility
- Video: What is Marginal Utility?
- Self Check: Defining Utility

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

126. Reading: Consumer Choices

"Eeny, Meeny, Miney, Moe"—Making Choices

Figure 6.1. Investment Choices. Higher education is generally viewed as a good investment, if one can afford it, regardless of the state of the economy. (Credit: modification of work by Jason Bache/ Flickr Creative Commons)

The Great Recession of 2008–2009 touched families around the globe. In many countries, workers found themselves out of a job. In developed countries, unemployment compensation provided a safety net, but families still saw a marked decrease in disposable income and had to make tough spending decisions. Of course, non-essential, discretionary spending was the first to go.

Even so, there was one particular category that saw a universal increase in spending world-wide during that time—an 18% uptick in the United States, specifically. You might guess that consumers began eating more meals at home, increasing spending at the grocery store. But the Bureau of Labor Statistics' Consumer Expenditure Survey, which tracks U.S. food spending over time, showed "real total food spending by U.S. households declined five percent between 2006 and 2009." So, it was not groceries. Just what product would people around the world demand more of during tough economic times, and more important, why?

That question leads us to this module's topic—analyzing how consumers make choices. For most consumers, using "eeny, meeny, miney, moe" is not how they make decisions; their decision-making processes have been educated far beyond a children's rhyme.

Introduction to Consumer Choices

Microeconomics seeks to understand the behavior of individual economic agents such as individuals and businesses. Economists believe that individuals' decisions, such as what goods and services to buy, can be analyzed as choices made within certain budget constraints. Generally, consumers are trying to get the most for their limited budget. In economic terms they are trying to maximize total utility, or satisfaction, given their budget constraint.

Everyone has their own personal tastes and preferences. The French say: *Chacun à son goût*, or "Each to his own taste." An old Latin saying states, *De gustibus non est disputandum* or "There's no disputing about taste." If people's decisions are based on their own tastes and personal preferences, then how can economists hope to analyze the choices consumers make?

An economic explanation for why people make different choices begins with accepting the proverbial wisdom that tastes are a matter of personal preference. But economists also believe that the choices people make are influenced by their incomes, by the prices of goods and services they consume, and by factors like where they live. This module introduces the economic theory of how consumers make choices about what to buy, how much to work, and how much to save. It will also illustrate how economic theory provides a tool to systematically look at the full range of possible consumption choices to predict how consumption responds to changes in prices or incomes.

127. Reading: Consumption Choices

Consumption Choices

Information on the ***consumption*** choices of Americans is available from the Consumer Expenditure Survey carried out by the U.S. Bureau of Labor Statistics. Table 6.1 shows spending patterns for the average U.S. household. The first row shows income and, after taxes and personal savings are subtracted, it shows that, in 2011, the average U.S. household spent $49,705 on consumption. The table then breaks down consumption into various categories. The average U.S. household spent roughly one-third of its consumption on shelter and other housing expenses, another one-third on food and vehicle expenses, and the rest on a variety of items, as shown. Of course, these patterns will vary for specific households by differing levels of family income, by geography, and by preferences.

Table 6.1. U.S. Consumption Choices in 2011	
Average U.S. household income before taxes	**$63,685**
Average annual expenditure	**$49,705**
Food at home	$3,838 (8%)
Food away from home	$2,620 (5%)
Housing	$16,803 (34%)
Apparel and services	$1,740 (4%)
Transportation	$8,293 (17%)
Healthcare	$3,313 (7%)
Entertainment	$2,572 (5%)
Education	$1,051 (2%)
Personal insurance and pensions	$5,424 (11%)
All else: alcohol, tobacco, reading, personal care, cash contributions, miscellaneous	$4,051 (8%)

TOTAL UTILITY AND DIMINISHING MARGINAL UTILITY

To understand how a household will make its choices, economists look at what consumers can afford, as shown in a ***budget constraint line***, and the ***total utility*** or satisfaction derived from those choices. In a budget constraint line, the quantity of one good is measured on the horizontal axis and the quantity of the other good is measured on the vertical axis. The budget constraint line shows the various combinations of two goods that are affordable given consumer income. Consider the situation of José, shown in Figure 6.2. José likes to collect T-shirts and watch movies.

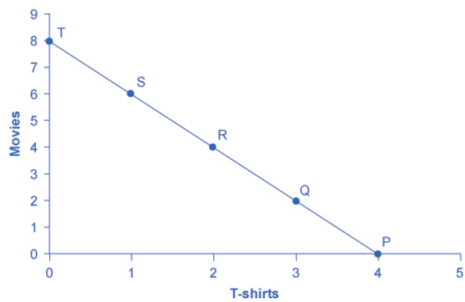


Figure 6.2. *A Choice between Consumption Goods. José has income of $56. Movies cost $7 and T-shirts cost $14. The points on the budget constraint line show the combinations of movies and T-shirts that are affordable.*

In Figure 6.2, the quantity of T-shirts is shown on the horizontal axis, while the quantity of movies is shown on the vertical axis. If José had unlimited income or goods were free, then he could consume without limit. But José, like all of us, faces a ***budget constraint***. José has a total of $56 to spend. The price of T-shirts is $14 and the price of movies is $7. Notice that the vertical intercept of the budget constraint line is at eight movies and zero T-shirts ($56/$7=8). The horizontal intercept of the budget constraint is four, where José spends of all of his money on T-shirts and no movies ($56/14=4). The slope of the budget constraint line is rise/run or −8/4=−2. The specific choices along the budget constraint line show the combinations of T-shirts and movies that are affordable.

José wishes to choose the combination that will provide him with the greatest ***utility***, which is the term economists use to describe a person's level of satisfaction or happiness with his or her choices.

Let's begin with an assumption, which will be discussed in more detail later, that José can measure his own utility with something called *utils*. (It is important to note that you cannot make comparisons between the utils of individuals; if one person gets 20 utils from a cup of coffee and another gets 10 utils, this does not mean than the first person gets more enjoyment from the coffee than the other or that they enjoy the coffee twice as much.) Table 6.2 shows how José's utility is connected with his consumption of T-shirts or movies. The first column of the table shows the quantity of T-shirts consumed. The second column shows the total utility, or total amount of satisfaction, that José receives from consuming that number of T-shirts. The most common pattern of total utility, as shown here, is that consuming additional goods leads to greater total utility, but at a decreasing rate. The third column shows ***marginal utility***, which is the additional utility provided by one additional unit of consumption. This equation for marginal utility is:

$$\text{MU} = \frac{\text{change in total utility}}{\text{change in quantity}}$$

Notice that marginal utility diminishes as additional units are consumed, which means that each subsequent unit of a good consumed provides less *additional* utility. For example, the first T-shirt José picks is his favorite and it gives him an addition of 22 utils. The fourth T-shirt is just something to wear when all his other clothes are in the wash and yields only 18 additional utils. This is an example of the law of ***diminishing marginal utility***, which holds that the additional utility decreases with each unit added.

The rest of Table 6.2 shows the quantity of movies that José attends, and his total and marginal utility from seeing each movie. Total utility follows the expected pattern: it increases as the number of movies seen rises. Marginal utility also follows the expected pattern: each additional movie brings a smaller gain in utility than the previous one. The first movie José attends is the one he wanted to see the most, and thus provides him with the highest level of utility or satisfaction. The fifth movie he attends is just to kill time. Notice that total utility is also the sum of the marginal utilities.

Table 6.2. Total and Marginal Utility

T-Shirts (Quantity)	Total Utility	Marginal Utility	Movies (Quantity)	Total Utility	Marginal Utility
1	22	22	**1**	16	16
2	43	21	**2**	31	15
3	63	20	**3**	45	14
4	81	18	**4**	58	13
5	97	16	**5**	70	12
6	111	14	**6**	81	11
7	123	12	**7**	91	10
8	133	10	**8**	100	9

Table 6.3 looks at each point on the budget constraint in Figure 6.2, and adds up José's total utility for five possible combinations of T-shirts and movies.

Table 6.3. Finding the Choice with the Highest Utility

Point	T-Shirts	Movies	Total Utility
P	4	0	81 + 0 = 81
Q	3	2	63 + 31 = 94
R	2	4	43 + 58 = 101
S	1	6	22 + 81 = 103
T	0	8	0 + 100 = 100

CALCULATING TOTAL UTILITY

Let's look at how José makes his decision in more detail:

Step 1. Observe that, at point Q (for example), José consumes three T-shirts and two movies.

Step 2. Look at Table 6.2. You can see from the fourth row/second column that three T-shirts are worth 63 utils. Similarly, the second row/fifth column shows that two movies are worth 31 utils.

Step 3. From this information, you can calculate that point Q has a total utility of 94 (63 + 31).

Step 4. You can repeat the same calculations for each point on Table 6.3, in which the total utility numbers are shown in the last column.

For José, the highest total utility for all possible combinations of goods occurs at point S, with a total utility of 103 from consuming one T-shirt and six movies.

128. Reading: Choosing with Marginal Utility

Choosing with Marginal Utility

Most people approach their utility-maximizing combination of choices in a step-by-step way. This step-by-step approach is based on looking at the tradeoffs, measured in terms of marginal utility, of consuming less of one good and more of another. You can think of this step-by-step approach as the “biggest bang for the buck” principle.

For example, say that José starts off thinking about spending all his money on T-shirts and choosing point P, which corresponds to four T-shirts and no movies, as illustrated in Figure 6.2.

The points on the graph show how a budget is affected by spending choices. Spending more money at the movies (y-axis) means that Jose' has less money to spend on T-shirts (x-axis).

Figure 6.2 *.A Choice between Consumption Goods. José has income of $56. Movies cost $7 and T-shirts cost $14. The points on the budget constraint line show the combinations of movies and T-shirts that are affordable.*

José chooses this starting point randomly; he has to start somewhere. Then he considers giving up the last T-shirt, the one that provides him the least marginal utility, and using the money he saves to buy two movies instead. Table 6.4 tracks the step-by-step series of decisions José needs to make (*Key*: T-shirts are $14, movies are $7, and income is $56).

Table 6.4. A Step-by-Step Approach to Maximizing Utility

Try	Which Has	Total Utility	Marginal Gain and Loss of Utility, Compared with Previous Choice	Conclusion
Choice 1: P	4 T-shirts and 0 movies	81 from 4 T-shirts + 0 from 0 movies = 81	–	–
Choice 2: Q	3 T-shirts and 2 movies	63 from 3 T-shirts + 31 from 0 movies = 94	Loss of 18 from 1 less T-shirt, but gain of 31 from 2 more movies, for a net utility gain of 13	Q is preferred over P
Choice 3: R	2 T-shirts and 4 movies	43 from 2 T-shirts + 58 from 4 movies = 101	Loss of 20 from 1 less T-shirt, but gain of 27 from two more movies for a net utility gain of 7	R is preferred over Q
Choice 4: S	1 T-shirt and 6 movies	22 from 1 T-shirt + 81 from 6 movies = 103	Loss of 21 from 1 less T-shirt, but gain of 23 from two more movies, for a net utility gain of 2	S is preferred over R
Choice 5: T	0 T-shirts and 8 movies	0 from 0 T-shirts + 100 from 8 movies = 100	Loss of 22 from 1 less T-shirt, but gain of 19 from two more movies, for a net utility loss of 3	S is preferred over T

DECISION MAKING BY COMPARING MARGINAL UTILITY

José could use the following thought process (if he thought in utils) to make his decision regarding how many T-shirts and movies to purchase:

Step 1. From Table 6.4, José can see that the marginal utility of the fourth T-shirt is 18. If José gives up the fourth T-shirt, then he loses 18 utils.

Step 2. Giving up the fourth T-shirt, however, frees up $14 (the price of a T-shirt), allowing José to buy the first two movies (at $7 each).

Step 3. José knows that the marginal utility of the first movie is 16 and the marginal utility of the second movie is 15. Thus, if José moves from point P to point Q, he gives up 18 utils (from the T-shirt), but gains 31 utils (from the movies).

Step 4. Gaining 31 utils and losing 18 utils is a net gain of 13. This is just another way of saying that the total utility at Q (94 according to the last column in Table 6.3) is 13 more than the total utility at P (81).

Step 5. So, for José, it makes sense to give up the fourth T-shirt in order to buy two movies.

José clearly prefers point Q to point P. Now repeat this step-by-step process of decision making with marginal utilities. José thinks about giving up the third T-shirt and surrendering a marginal utility of 20, in exchange for purchasing two more movies that promise a combined marginal utility of 27. José prefers point R to point Q. What if José thinks about going beyond R to point S? Giving up the second T-shirt means a marginal utility loss of 21, and the marginal utility gain from the fifth and sixth movies would combine to make a marginal utility gain of 23, so José prefers point S to R.

However, if José seeks to go beyond point S to point T, he finds that the loss of marginal utility from giving up the first T-shirt is 22, while the marginal utility gain from the last two movies is only a total of 19. If José were to choose point T, his utility would fall to 100. Through these stages of thinking about marginal tradeoffs, José again concludes that S, with one T-shirt and six movies, is the choice that will provide him with the highest level of total utility. This step-by-step approach will reach the same conclusion regardless of José's starting point.

Another way to look at this is by focusing on satisfaction per dollar. M*arginal utility per dollar* is the amount of additional utility José receives given the price of the product.

marginal utility per dollar = marginal utility/price

For José's T-shirts and movies, the marginal utility per dollar is shown in Table 6.5.

Table 6.5 Marginal Utility per Dollar

Quantity of T-Shirts	Total Utility	Marginal Utility	Marginal Utility per Dollar	Quantity of Movies	Total Utility	Marginal Utility	Marginal Utility per Dollar
1	**22**	**22**	**22/$14=1.6**	1	16	16	16/$7=2.3
2	43	21	21/$14=1.5	2	31	15	15/$7=2.14
3	63	20	20/$14=1.4	3	45	14	14/$7=2
4	81	18	18/$14=1.3	4	58	13	13/$7=1.9
5	97	16	16/$14=1.1	5	70	12	12/$7=1.7
6	111	14	14/$14=1	**6**	**81**	**11**	**11/$7=1.6**
7	123	12	12/$14=1.2	7	91	10	10/$7=1.4

José's first purchase will be a movie. Why? Because it gives him the highest marginal utility per dollar and it is affordable. José will continue to purchase the good which gives him the highest marginal utility per dollar until he exhausts the budget. José will keep purchasing movies because they give him a greater "bang for the buck" until the sixth movie is equivalent to a T-shirt purchase. José can afford to purchase that T-shirt. So José will choose to purchase six movies and one T-shirt.

Self Check: Defining Utility

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the three Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/634

129. Video: What is Marginal Utility?

This tutorial on marginal utility will walk you through the thought process of purchasing a burger at a store. The benefit you get from the burger is considered your marginal utility. You like it so much, you buy a second burger and enjoy it once more. The extra satisfaction you acquired only from the second burger is known as the marginal utility, whereas the total enjoyment from both burgers is considered the total utility.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=154#oembed-1

130. Outcome: Marginal Utility

What you'll learn to do: describe and calculate the concept of marginal utility

If you look at Facebook for five fewer minutes a day, will you really be more productive? Is it worth it to clock into work ten minutes early, or would it be best to spend that time with your significant other? Is it worth it to spend five dollars on a dessert when you already feel a little bit full? In this section, you will examine choices made at the margin, or the decisions you make to do a little more or a little less of something.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Marginal Utility
- Video: Law of Diminishing Marginal Utility
- Self Check: Marginal Utility

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

131. Reading: Marginal Utility

Choices Are Made at the Margin

Economists argue that most choices are made "at the margin." The margin is the current level of an activity. Think of it as the edge from which a choice is to be made. A choice at the margin is a decision to do a little more or a little less of something.

Assessing choices at the margin can lead to extremely useful insights. Consider, for example, the problem of curtailing water consumption when the amount of water available falls short of the amount people now use. Economists argue that one way to induce people to conserve water is to raise its price. A common response to this recommendation is that a higher price would have no effect on water consumption, because water is a necessity. Many people assert that prices do not affect water consumption because people "need" water.

But choices in water consumption, like virtually all choices, are made at the margin. Individuals do not make choices about whether they should or should not consume water. Rather, they decide whether to consume a little more or a little less water. Household water consumption in the United States totals about 105 gallons per person per day. Think of that starting point as the edge from which a choice at the margin in water consumption is made. Could a higher price cause you to use less water brushing your teeth, take shorter showers, or water your lawn less? Could a higher price cause people to reduce their use, say, to 104 gallons per person per day? To 103? When we examine the choice to consume water at the margin, the notion that a higher price would reduce consumption seems much more plausible. Prices affect our consumption of water because choices in water consumption, like other choices, are made at the margin.

The elements of opportunity cost, maximization, and choices at the margin can be found in each of two broad areas of economic analysis: microeconomics and macroeconomics. Your economics course, for example, may be designated as a “micro” or as a “macro” course. We will look at these two areas of economic thought in the next section.

Marginal utility is based on the notion that individuals rarely face all-or-nothing decisions, that most of the time they are considering a little more or a little less of something when allocating their budget, time or other scarce resources. So a student might ask, “how much better will I do on an exam if I study for one more hour?” The answer could be called the marginal grade improvement. Economists use a number of marginal concepts. The marginal utility of a third slice of pizza is the change in satisfaction one gets when eating the third slice instead of stopping with two. The marginal cost of one more unit of output a firm produces is the amount that total cost increases when the firm produces one more unit of output.

The general formula for computing a marginal item is the change in the outcome divided by the change in the number of inputs used to produce that outcome. For example, if two more hours of work yields an additional $20 in wages, the marginal wage earned is $20/2 hours = $10 per hour.

Law of Diminishing Marginal Utility

Watch this lecture video clip to learn more about why when we consume even more of a good, the marginal benefit of that good decreases with each additional good.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=156#oembed-1

Self Check: Marginal Utility

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/636

132. Outcome: The Utility Maximizing Rule

What you'll learn to do: explain how consumers maximize total utility within a given income using the Utility Maximizing Rule

In this section, you'll learn how exactly to measure when you are getting "more bang for your buck."

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: A Tool for Maximizing Utility
- Video: The Optimal Purchase Rule
- Simulation: Maximizing Utility
- Self Check: The Utility Maximizing Rule

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

133. Reading: A Tool for Maximizing Utility

A Tool for Maximizing Utility

This process of decision making suggests a rule to follow when ***maximizing utility***. Since the price of T-shirts is twice as high as the price of movies, to maximize utility the last T-shirt chosen needs to provide exactly twice the marginal utility (MU) of the last movie. If the last T-shirt provides less than twice the marginal utility of the last movie, then the T-shirt is providing less "bang for the buck" (i.e., marginal utility per dollar spent) than if the same money were spent on movies. If this is so, José should trade the T-shirt for more movies to increase his total utility. Marginal utility per dollar measures the additional utility that José will enjoy given what he has to pay for the good.

Review José's T-shirts and movies marginal utility per dollar Table again.

Table 6.3. Marginal Utility per Dollar

Quantity of T-Shirts	Total Utility	Marginal Utility	Marginal Utility per Dollar	Quantity of Movies	Total Utility	Marginal Utility	Marginal Utility per Dollar
1	**22**	**22**	**22/$14=1.6**	1	16	16	16/$7=2.3
2	43	21	21/$14=1.5	2	31	15	15/$7=2.14
3	63	20	20/$14=1.4	3	45	14	14/$7=2
4	81	18	18/$14=1.3	4	58	13	13/$7=1.9
5	97	16	16/$14=1.1	5	70	12	12/$7=1.7
6	111	14	14/$14=1	**6**	**81**	**11**	**11/$7=1.6**
7	123	12	12/$14=1.2	7	91	10	10/$7=1.4

If the last T-shirt provides more than twice the marginal utility of the last movie, then the T-shirt is providing more "bang for the buck" or marginal utility per dollar, than if the money were spent on movies. As a result, José should buy more T-shirts. Notice that at José's optimal choice of point S, the marginal utility from the first T-shirt, of 22 is exactly twice the marginal utility of the sixth movie, which is 11. At this choice, the marginal utility per dollar is the same for both goods. This is a tell-tale signal that José has found the point with highest total utility.

This argument can be written as a general rule: the utility-maximizing choice between consumption goods occurs where the marginal utility per dollar is the same for both goods.

$$\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$$

A sensible economizer will pay twice as much for something only if, in the marginal comparison, the item confers twice as much utility. Notice that the formula for the table above is

$$\frac{22}{14} = \frac{11}{7}$$

$$1.6 = 1.6$$

The following feature provides step-by-step guidance for this concept of utility-maximizing choices.

Maximizing Utility

The general rule, $\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$, means that the last dollar spent on each good provides exactly the same marginal utility. So:

Step 1. If we traded a dollar more of movies for a dollar more of T-shirts, the marginal utility gained from T-shirts would exactly offset the marginal utility lost from fewer movies. In other words, the net gain would be zero.

Step 2. Products, however, usually cost more than a dollar, so we cannot trade a dollar's worth of movies. The best we can do is trade two movies for another T-shirt, since in this example T-shirts cost twice what a movie does.

Step 3. If we trade two movies for one T-shirt, we would end up at point R (two T-shirts and four movies).

Step 4. Choice 4 in Table 6.4 shows that if we move to point S, we would lose 21 utils from one less T-shirt, but gain 23 utils from two more movies, so we would end up with more total utility at point S.

Table 6.4. A Step-by-Step Approach to Maximizing Utility

Try	Which Has	Total Utility	Marginal Gain and Loss of Utility, Compared with Previous Choice	Conclusion
Choice 1: P	4 T-shirts and 0 movies	81 from 4 T-shirts + 0 from 0 movies = 81	–	–
Choice 2: Q	3 T-shirts and 2 movies	63 from 3 T-shirts + 31 from 0 movies = 94	Loss of 18 from 1 less T-shirt, but gain of 31 from 2 more movies, for a net utility gain of 13	Q is preferred over P
Choice 3: R	2 T-shirts and 4 movies	43 from 2 T-shirts + 58 from 4 movies = 101	Loss of 20 from 1 less T-shirt, but gain of 27 from two more movies for a net utility gain of 7	R is preferred over Q
Choice 4: S	1 T-shirt and 6 movies	22 from 1 T-shirt + 81 from 6 movies = 103	Loss of 21 from 1 less T-shirt, but gain of 23 from two more movies, for a net utility gain of 2	S is preferred over R
Choice 5: T	0 T-shirts and 8 movies	0 from 0 T-shirts + 100 from 8 movies = 100	Loss of 22 from 1 less T-shirt, but gain of 19 from two more movies, for a net utility loss of 3	S is preferred over T

In short, the general rule shows us the utility-maximizing choice.

There is another, equivalent way to think about this. The general rule can also be expressed as *the ratio of the prices of the two goods should be equal to the ratio of the marginal utilities.* When the price of good 1 is divided by the price of good 2, at the utility-maximizing point this will equal the marginal utility of good 1 divided by the marginal utility of good 2. This rule, known as the *consumer equilibrium*, can be written in algebraic form:

$$\frac{P_1}{P2} = \frac{MU_1}{MU_2}$$

Along the budget constraint, the total price of the two goods remains the same, so the ratio of the prices does not change. However, the marginal utility of the two goods changes with the quantities consumed. At the optimal choice of one T-shirt and six movies, point S, the ratio of marginal utility to price for T-shirts (22:14) matches the ratio of marginal utility to price for movies (of 11:7).

MEASURING UTILITY WITH NUMBERS

This discussion of utility started off with an assumption that it is possible to place numerical values on utility, an assumption that may seem questionable. You can buy a thermometer for measuring temperature at the hardware store, but what store sells an "utilimometer" for measuring utility? However, while measuring utility with numbers is a convenient assumption to clarify the explanation, the key assumption is not that utility can be measured by an outside party, but only that individuals can decide which of two alternatives they prefer.

To understand this point, think back to the step-by-step process of finding the choice with highest total utility by comparing the marginal utility that is gained and lost from different choices along the budget constraint. As José compares each choice along his budget constraint to the previous choice, what matters is not the specific numbers that he places on his utility—or whether he uses any numbers at all—but only that he personally can identify which choices he prefers.

In this way, the step-by-step process of choosing the highest level of utility resembles rather closely how many people make consumption decisions. We think about what will make us the happiest; we think about what things cost; we think about buying a little more of one item and giving up a little of something else; we choose what provides us with the greatest level of satisfaction. The vocabulary of comparing the points along a budget constraint and total and marginal utility is just a set of tools for discussing this everyday process in a clear and specific manner. It is welcome news that specific utility numbers are not central to the argument, since a good utilimometer is hard to find. Do not worry—while we cannot measure utils, by the end of the next module, we will have transformed our analysis into something we can measure—demand.

134. Video: The Optimal Purchase Rule

This video will take you through an example to understand why the optimal purchasing point is at the point where the price is equal to marginal utility.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=159#oembed-1

135. Simulation: Maximizing Utility

Try It

Play the simulation below multiple times to see how different choices lead to different outcomes. All simulations allow unlimited attempts so that you can gain experience applying the concepts.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=160

136. Self Check: The Utility Maximizing Rule

Check Your Understanding

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/637

137. Outcome: How Utility Changes

What you'll learn to do: explain how consumer's utility changes when income or prices change

Imagine you were to create a pie chart of your college budget and spending. Where does most of your money go? Is it spent on housing? Tuition? Maybe on a car payment or food? How much money goes towards savings, entertainment, or a night out with friends? In this outcome, you'll see how economists interpret these consumer choices and learn how to analyze utility considering a person's budget constraints.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: The Foundations of Demand Curve
- Reading: Income Changes and Consumption Choices
- Self Check: Consumption Choices

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

138. Reading: The Foundations of Demand Curve

Remember that a demand curve shows the relationship between price and quantity demanded. While demand curves will appear somewhat different for each product – they may appear relatively steep or flat, straight or curved – nearly all demand curves slope down from left to right. So demand curves embody the law of demand: as the price increases, the quantity demanded decreases, and conversely, as the price decreases, the quantity demanded increases.

The Foundations of Demand Curves

Changes in the price of a good lead the budget constraint to shift. A shift in the budget constraint means that when individuals are seeking their highest utility, the quantity that is demanded of that good will change. In this way, the logical foundations of demand curves—which show a connection between prices and quantity demanded—are based on the underlying idea of individuals seeking utility. Figure 6.5 (a) shows a budget constraint with a choice between housing and "everything else." (Putting "everything else" on the vertical axis can be a useful approach in some cases, especially when the focus of the analysis is on one particular good.) The preferred choice on the original budget constraint that provides the highest possible utility is labeled M0. The other three budget constraints represent successively higher prices for housing of P1, P2, and P3. As the budget constraint rotates in, and in, and in again, the utility-maximizing choices are labeled M1, M2, and M3, and the quantity demanded of housing falls from Q0 to Q1 to Q2 to Q3.

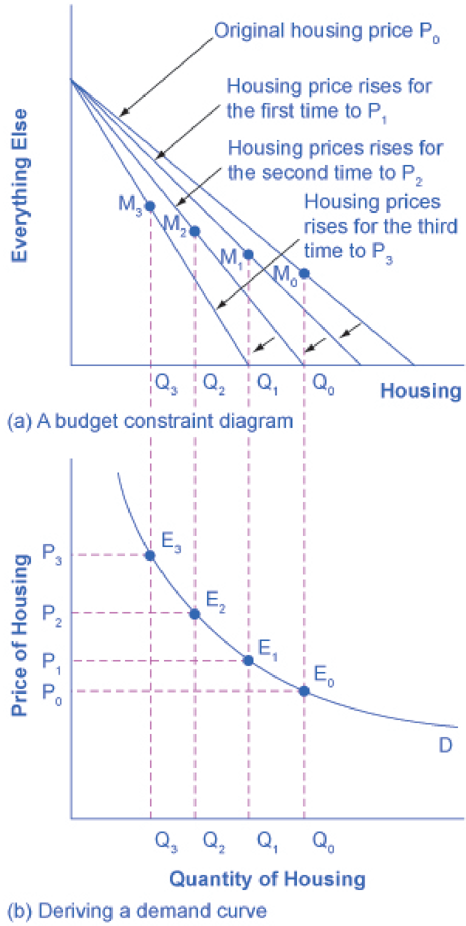


***Figure 6.5.** The Foundations of a Demand Curve: An Example of Housing. (a) As the price increases from P0 to P1 to P2 to P3, the budget constraint on the upper part of the diagram shifts to the left. The utility-maximizing choice changes from M0 to M1 to M2 to M3. As a result, the quantity demanded of housing shifts from Q0 to Q1 to Q2 to Q3, ceteris paribus. (b) The demand curve graphs each combination of the price of housing and the quantity of housing demanded, ceteris paribus. Indeed, the quantities of housing are the same at the points on both (a) and*

(b). Thus, the original price of housing (P0) and the original quantity of housing (Q0) appear on the demand curve as point E0. The higher price of housing (P1) and the corresponding lower quantity demanded of housing (Q1) appear on the demand curve as point E1.

So, as the price of housing rises, the budget constraint shifts to the left, and the quantity consumed of housing falls, *ceteris paribus* (meaning, with all other things being the same). This relationship—the price of housing rising from P0 to P1 to P2 to P3, while the quantity of housing demanded falls from Q0 to Q1 to Q2 to Q3—is graphed on the demand curve in Figure 6.5 (b). Indeed, the vertical dashed lines stretching between the top and bottom of Figure 6.5 show that the quantity of housing demanded at each point is the same in both (a) and (b). The shape of a demand curve is ultimately determined by the underlying choices about maximizing utility subject to a budget constraint. And while economists may not be able to measure "utils," they can certainly measure price and quantity demanded.

Applications in Government and Business

The budget constraint framework for making utility-maximizing choices offers a reminder that people can react to a change in price or income in a range of different ways. For example, in the winter months of 2005, costs for heating homes increased significantly in many parts of the country as prices for natural gas and electricity soared, due in large part to the disruption caused by Hurricanes Katrina and Rita. Some people reacted by reducing the quantity demanded of energy; for example, by turning down the thermostats in their homes by a few degrees and wearing a heavier sweater inside. Even so, many home heating bills rose, so people adjusted their consumption in other ways, too. As you learned in the module on Elasticity, the short run demand for home heating is generally inelastic. Each household cut back on what it valued least on the margin; for some it might have been some dinners out, or a vacation, or postponing buying a new refrigerator or a new car. Indeed, sharply higher energy prices can have effects beyond the energy market, leading to a widespread reduction in purchasing throughout the rest of the economy.

A similar issue arises when the government imposes taxes on certain products, like it does on gasoline, cigarettes, and alcohol. Say that a tax on alcohol leads to a higher price at the liquor store, the higher price of alcohol causes the budget constraint to pivot left, and consumption of alcoholic beverages is likely to decrease. However, people may also react to the higher price of alcoholic beverages by cutting back on other purchases. For example, they might cut back on snacks at restaurants like chicken wings and nachos. It would be unwise to assume that the liquor industry is the only one affected by the tax on alcoholic beverages.

Does it make a difference who controls household income?

In the mid-1970s, the United Kingdom made an interesting policy change in its "child allowance" policy. This program provides a fixed amount of money per child to every family, regardless of family income. Traditionally, the child allowance had been distributed to families by withholding less in taxes from the paycheck of the family wage earner—typically the father in this time period. The new policy instead provided the child allowance as a cash payment to the mother. As a result of this change, households have the same level of income and face the same prices in the market, but the money is more likely to be in the purse of the mother than in the wallet of the father.

Should this change in policy alter household consumption patterns? Basic models of consumption decisions, of the sort examined in this module, assume that it does not matter whether the mother or the father receives the money, because both parents seek to maximize the utility of the family as a whole. In effect, this model assumes that everyone in the family has the same preferences.

In reality, the share of income controlled by the father or the mother does affect what the household consumes. When the mother controls a larger share of family income a number of studies, in the United Kingdom and in a wide variety of other countries, have found that the family tends to spend more on restaurant meals, child care, and women's clothing, and less on alcohol and tobacco. As the mother controls a larger share of household resources, children's health improves, too. These findings suggest that when providing assistance to poor families, in high-income countries and low-income countries alike, the monetary amount of assistance is not all that matters: it also matters which member of the family actually receives the money.

The budget constraint framework serves as a constant reminder to think about the full range of effects that can arise from changes in income or price, not just effects on the one product that might seem most immediately affected.

139. Reading: Income Changes and Consumption Choices

How Changes in Income Affect Consumer Choices

Just as utility and marginal utility can be used to discuss making consumer choices along a budget constraint, these ideas can also be used to think about how consumer choices change when the budget constraint shifts in response to changes in income or price. Indeed, because the budget constraint framework can be used to analyze how quantities demanded change because of price movements, the budget constraint model can illustrate the underlying logic behind demand curves.

Let's begin with a concrete example illustrating how changes in income level affect consumer choices. Figure 6.3 shows a ***budget constraint*** that represents Kimberly's choice between concert tickets at $50 each and getting away overnight to a bed-and-breakfast for $200 per night. Kimberly has $1,000 per year to spend between these two choices. After thinking about her total utility and marginal utility and applying the decision rule that the ratio of the marginal utilities to the prices should be equal between the two products, Kimberly chooses point M, with eight concerts and three overnight getaways as her utility-maximizing choice.

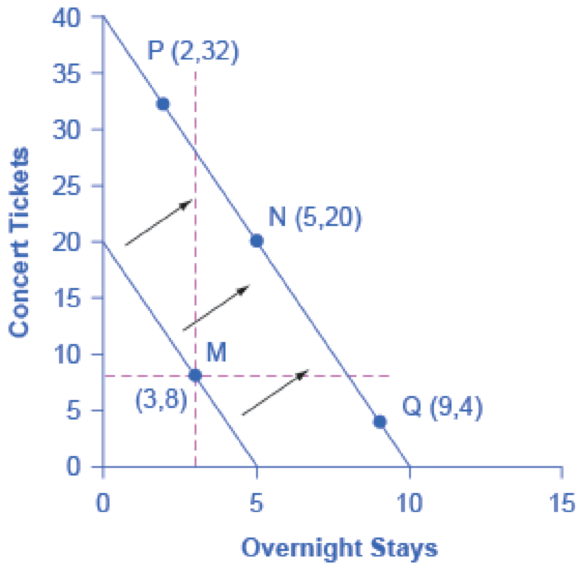


Figure 6.3. *How a Change in Income Affects Consumption Choices. The utility-maximizing choice on the original budget constraint is M. The dashed horizontal and vertical lines extending through point M allow you to see at a glance whether the quantity consumed of goods on the new budget constraint is higher or lower than on the original budget constraint. On the new budget constraint, a choice like N will be made if both goods are normal goods. If overnight stays is an inferior good, a choice like P will be made. If concert tickets are an inferior good, a choice like Q will be made.*

Now, assume that the income Kimberly has to spend on these two items rises to $2,000 per year, causing her budget constraint to shift out to the right. How does this rise in income alter her utility-maximizing choice? Kimberly will again consider the utility and marginal utility that she receives from concert tickets and overnight getaways and seek her ***utility-maximizing choice*** on the new budget line. But how will her new choice relate to her original choice?

The possible choices along the new budget constraint can be divided into three groups, which are divided up by the dashed horizontal and vertical lines that pass through the original choice M in the figure. All choices on the upper left of the new budget constraint that are to the left of the vertical dashed line, like choice P with two overnight stays and 32 concert tickets, involve less of the good on the horizontal axis but much more of the good on the vertical axis. All choices to the right of the vertical dashed line and above the horizontal dashed line—like choice N with five overnight getaways and 20 concert tickets—have more consumption of both goods. Finally, all choices that are to the right of the vertical dashed line but below the horizontal dashed line, like choice Q with four concerts and nine overnight getaways, involve less of the good on the vertical axis but much more of the good on the horizontal axis.

All of these choices are theoretically possible, depending on Kimberly's personal preferences as expressed through the total and marginal utility she would receive from consuming these two goods. When income rises, the most common reaction is to purchase more of both goods, like choice N, which is to the upper right relative to Kimberly's original choice M, although exactly how much more of each good will vary according to personal taste. Conversely, when income falls, the most typical reaction is to purchase less of both goods. As defined in the module on Demand and Supply and again in the module on Elasticity, goods and services are called *normal goods* when a rise in income leads to a rise in the quantity consumed of that good and a fall in income leads to a fall in quantity consumed.

However, depending on Kimberly's preferences, a rise in income could cause consumption of one good to increase while consumption of the other good declines. A choice like P means that a rise in income caused her quantity consumed of overnight stays to decline, while a choice like Q would mean that a rise in income caused her quantity of concerts to decline. Goods where demand declines as income rises (or conversely, where the demand rises as income falls) are called "inferior goods." An ***inferior good*** occurs when people trim back on a good as income rises, because they can now afford the more expensive choices that they prefer. For example, a higher-income household might eat fewer hamburgers or be less likely to buy a used car, and instead eat more steak and buy a new car.

How Price Changes Affect Consumer Choices

For analyzing the possible effect of a change in price on consumption, let's again use a concrete example. Figure 6.4 represents the consumer choice of Sergei, who chooses between purchasing baseball bats and cameras. A price increase for baseball bats would have no effect on the ability to purchase cameras, but it would reduce the number of bats Sergei could afford to buy. Thus a price increase for baseball bats, the good on the horizontal axis, causes the budget constraint to rotate inward, as if on a hinge, from the vertical axis. As in the previous section, the point labeled M represents the originally preferred point on the original budget constraint, which Sergei has chosen after contemplating his total utility and marginal utility and the tradeoffs involved along the budget constraint. In this example, the units along the horizontal and vertical axes are not numbered, so the discussion must focus on whether more or less of certain goods will be consumed, not on numerical amounts.

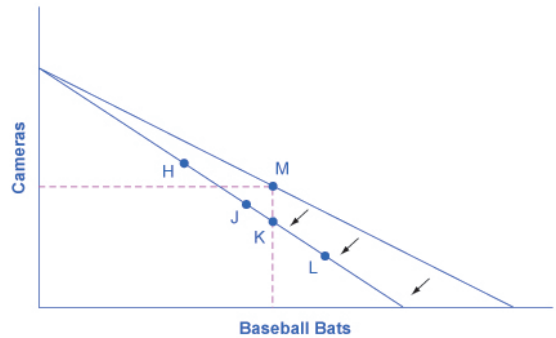


Figure 6.4. *How a Change in Price Affects Consumption Choices. The original utility-maximizing choice is M. When the price rises, the budget constraint shifts in to the left. The dashed lines make it possible to see at a glance whether the new consumption choice involves less of both goods, or less of one good and more of the other. The new possible choices would be fewer baseball bats and more cameras, like point H, or less of both goods, as at point J. Choice K would mean that the higher price of bats led to exactly the same quantity of bats being consumed, but fewer cameras. Choices like L are ruled out as theoretically possible but highly unlikely in the real world, because they would mean that a higher price for baseball bats means a greater quantity consumed of baseball bats.*

After the price increase, Sergei will make a choice along the new budget constraint. Again, his choices can be divided into three segments by the dashed vertical and horizontal lines. In the upper left portion of the new budget constraint, at a choice like H, Sergei consumes more cameras and fewer bats. In the central portion of the new budget constraint, at a choice like J, he consumes less of both goods. At the right-hand end, at a choice like L, he consumes more bats but fewer cameras.

The typical response to higher prices is that a person chooses to consume less of the product with the higher price. This occurs for two reasons, and both effects can occur simultaneously. The ***substitution effect*** occurs when a price changes and consumers have an incentive to consume less of the good with a relatively higher price and more of the good with a relatively lower price. The ***income effec***t is that a higher price means, in effect, the buying power of income has been reduced (even though actual income has not changed), which leads to buying less of the good (when the good is normal). In this example, the higher price for baseball bats would cause Sergei to buy fewer bats for both reasons. Exactly how much will a higher price for bats cause Sergei consumption of bats to fall? Figure 6.4 suggests a range of possibilities. Sergei might react to a higher price for baseball bats by purchasing the same quantity of bats, but cutting his consumption of cameras. This choice is the point K on the new budget constraint, straight below the original choice M. Alternatively, Sergei might react by dramatically reducing his purchases of bats and instead buy more cameras.

The key is that it would be imprudent to assume that a change in the price of baseball bats will only or primarily affect the good whose price is changed, while the quantity consumed of other goods remains the same. Since Sergei purchases all his products out of the same budget, a change in the price of one good can also have a range of effects, either positive or negative, on the quantity consumed of other goods.

In short, a higher price typically causes reduced consumption of the good in question, but it can affect the consumption of other goods as well.

LINK IT UP

Read this article about the potential of variable prices in vending machines.

Watch this video below to see an example of the relationship between price, demand, and marginal utility.

One or more interactive elements has been excluded *from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=164#oembed-1*

Self Check: How Utility Changes

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/638

140. Outcome: Behavioral Economics

What you'll learn to do: describe the behavioral economics approach to understanding decision making

If a friendly stranger handed you $20, how would you feel? What if you lost $20 the next week? Do you think that you'd feel sad about it, or would just be happy that you broke even? In this outcome, you'll see that most people would still be sad about the $20 loss. Economic decisions are not always logical. Read on to find out why!

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Behavioral Economics: An Alternative Viewpoint
- Video: Behavioral Economics
- Self Check: Behavioral Economics

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

141. Reading: Behavioral Economics: An Alternative Viewpoint

Behavioral Economics: An Alternative Viewpoint

As we know, people sometimes make decisions that seem "irrational" and not in their own best interest. People's decisions can seem inconsistent from one day to the next and they even deliberately ignore ways to save money or time. The traditional economic models assume rationality, which means that people take all available information and make consistent and informed decisions that are in their best interest. (In fact, economics professors often delight in pointing out so-called "irrational behavior" each semester to their new students, and present economics as a way to become more rational.)

But a new group of economists, known as behavioral economists, argue that the traditional method leaves out something important: people's state of mind. For example, one can think differently about money if one is feeling revenge, optimism, or loss. These are not necessarily irrational states of mind, but part of a range of emotions that can affect anyone on a given day. And what's more, actions under these conditions are indeed predictable, if the underlying environment is better understood. So, behavioral economics seeks to enrich the understanding of decision-making by integrating the insights of psychology into economics. It does this by investigating how given dollar amounts can mean different things to individuals depending on the situation. This can lead to decisions that appear outwardly inconsistent, or irrational, to the outside observer.

The way the mind works, according to this view, may seem inconsistent to traditional economists but is actually far more complex than an unemotional cost-benefit adding machine. For example, a traditional economist would say that if you lost a $10 bill today, and also got an extra $10 in your paycheck, you should feel perfectly neutral. After all, –$10 + $10 = $0. You are the same financially as you were before. However, behavioral economists have done research that shows many people will feel some negative emotion—anger, frustration, and so forth—after those two things happen. We tend to focus more on the loss than the gain. This is known as loss aversion, where a $1 loss pains us 2.25 times more than a $1 gain helps us, according to the economists Daniel Kahneman and Amos Tversky in a famous 1979 article in the journal *Econometrica*. This insight has implications for investing, as people tend to "overplay" the stock market by reacting more to losses than to gains. Indeed, this behavior looks irrational to traditional economists, but is consistent once we understand better how the mind works, these economists argue.

Irrational Consumer Behavior

Traditional economists also assume human beings have complete self-control. But, for instance, people will buy cigarettes by the pack instead of the carton even though the carton saves them money, to keep usage down. They purchase locks for their refrigerators and overpay on taxes to force themselves to save. In other words, we protect ourselves from our worst temptations but pay a price to do so. One way behavioral economists are responding to this is by setting up ways for people to keep themselves free of these temptations. This includes what are called "nudges" toward more rational behavior rather than mandatory regulations from government. For example, up to 20 percent of new employees do not enroll in retirement savings plans immediately, because of procrastination or feeling overwhelmed by the different choices.

Some companies are now moving to a new system, where employees are automatically enrolled unless they "opt out." Almost no-one opts out in this program and employees begin saving at the early years, which are most critical for retirement.

Another area that seems illogical is the idea of mental accounting, or putting dollars in different mental categories where they take different values. Economists typically consider dollars to be fungible, or having equal value to the individual, regardless of the situation.

You might, for instance, think of the $25 you found in the street differently from the $25 you earned from three hours working in a fast food restaurant. The street money might well be treated as "mad money" with little rational regard to getting the best value. This is in one sense strange, since it is still equivalent to three hours of hard work in the restaurant. Yet the "easy come-easy go" mentality replaces the rational economizer because of the situation, or context, in which the money was attained.

In another example of mental accounting that seems inconsistent to a traditional economist, a person could carry a credit card debt of $1,000 that has a 15% yearly interest cost, and simultaneously have a $2,000 savings account that pays only 2% per year. That means she pays $150 a year to the credit card company, while collecting only $40 annually in bank interest, so she loses $130 a year. That doesn't seem wise.

The "rational" decision would be to pay off the debt, since a $1,000 savings account with $0 in debt is the equivalent net worth, and she would now net $20 per year. But curiously, it is not uncommon for people to ignore this advice, since they will treat a loss to their savings account as higher than the benefit of paying off their credit card. The dollars are not being treated as fungible so it looks irrational to traditional economists.

Which view is right, the behavioral economists' or the traditional view? Both have their advantages, but behavioral economists have at least shed a light on trying to describe and explain behavior that has historically been dismissed as irrational. If most of us are engaged in some “irrational behavior,” perhaps there are deeper underlying reasons for this behavior in the first place.

“EENY, MEENY, MINEY, MOE”—MAKING CHOICES

In what category did consumers worldwide increase their spending during the recession? Higher education. According to the United Nations Educational, Scientific, and Cultural Organization (UNESCO), enrollment in colleges and universities rose one-third in China and almost two-thirds in Saudi Arabia, nearly doubled in Pakistan, tripled in Uganda, and surged by three million—18 percent—in the United States. Why were consumers willing to spend on education during lean times? Both individuals and countries view higher education as the way to prosperity. Many feel that increased earnings are a significant benefit of attending college.

Bureau of Labor Statistics data from May 2012 supports this view, as shown in Figure 3. They show a positive correlation between earnings and education. The data also indicate that unemployment rates fall with higher levels of education and training.

The graph shows the unemployment rate and median weekly earnings in 2012 for various levels of education. People with professional degrees made around $1,735 a week and suffered a 2.1% unemployment rate. People with doctoral degrees made around $1,624 a week and suffered a 2.5% unemployment rate. People with Master's degrees made around $1,300 a week and suffered a 3.5% unemployment rate. People with Bachelor's degrees made around $1,066 a week and suffered a 4.5% unemployment rate. People with Associate's degrees made around $785 a week and suffered a 6.2% unemployment rate. People with some college, no degree made around $727 a week and suffered a 7.7% unemployment rate. People with a high school diploma made around $652 a week and suffered an 8.3% unemployment rate. People with less than a high school diploma made around $471 a week and suffered a 12.4% unemployment rate.

***Figure 3. The Impact of Education on Earnings and Unemployment Rates, 2012**. Those with the highest degrees in 2012 had substantially lower unemployment rates whereas those with the least formal education suffered from the highest unemployment rates. The national median average weekly income was $815, and the nation unemployment average in 2012 was 6.8%. (Source: Bureau of Labor Statistics, May 22, 2013)*

Behavioral Economics Video

Although economists would love to assume that all people think rationally about their financial choices, events like the financial crisis led behavioral economists to look at how people actually make decisions in financial trading, instead of how they say they do, or even how they should. This video explains how human psychology played a large role in the financial crisis of 2008.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=166#oembed-1

Self Check: Behavioral Economics

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/639

142. Putting It Together: Utility

Summary

"*Budget*" *by Tax Credits, CC-BY.*

The goal of this module was to explain how consumer behavior shapes the demand curve with respect to utility and loss. You learned how to:

- Define the concept of utility, and to be able to differentiate between marginal and total utility

- Explain how consumers maximize total utility within a given income or budget using the Utility Maximizing Rule
- Explain how consumer's utility changes when income or prices change
- Describe the behavioral economics approach to understanding decision-making

Examples

The key underlying principle in this module was getting the biggest bang for the buck. This principle will be used over and over again in different contexts in this text, and the results will sometimes surprise you.

Let's return to the GPA example in the "Why it Matters" Feature. To keep it simple, let us suppose you are only taking two courses, an "easy" course, and a "hard" course, where the difficulty is defined by how much time and effort it takes to earn a given grade. Is Principles of Microeconomics an easy course or a hard one for you? I don't know. You can decide which of your courses falls into each category. Many people think that it makes the most sense to spend the most time and effort on the hard course. After all, a hard course requires more time to learn, right? That's true, but if you think like an economist, you'll see that to maximize your GPA, given a limited amount of study time, it makes more sense to start with the course where your study time will have the most impact on your grades, the biggest bang for the buck. In other words, you should start with the easy course and quite possibly spend more time on it, until to assure yourself of an A.

This is no different than choosing to spend more of your budget on the product that gives you the most marginal utility per dollar spent. It both cases, you make the most of your scarce resource, budget dollars in the consumption case and study hours in the GPA case.

As you proceed through the rest of this text, look for opportunities to apply the biggest bang for the buck principle.

143. Glossary: Utility

backward-bending supply curve for labor the situation when high-wage people can earn so much that they respond to a still-higher wage by working fewer hours

behavioral economics a branch of economics that seeks to enrich the understanding of decision-making by integrating the insights of psychology and by investigating how given dollar amounts can mean different things to individuals depending on the situation.

budget constraint line shows the possible combinations of two goods that are affordable given a consumer's limited income

consumer equilibrium when the ratio of the prices of goods is equal to the ratio of the marginal utilities (point at which the consumer can get the most satisfaction)

diminishing marginal utility the common pattern that each marginal unit of a good consumed provides less of an addition to utility than the previous unit

fungible the idea that units of a good, such as dollars, ounces of gold, or barrels of oil are capable of mutual substitution with each other and carry equal value to the individual.

income effect a higher price means that, in effect, the buying power of income has been reduced, even though actual income has not changed; always happens simultaneously with a substitution effect

marginal utility per dollar the additional satisfaction gained from purchasing a good given the price of the product; MU/Price

marginal utility the additional utility provided by one additional unit of consumption

substitution effect when a price changes, consumers have an incentive to consume less of the good with a relatively higher price and more of the good with a relatively lower price; always happens simultaneously with an income effect

total utility satisfaction derived from consumer choices

144. Discussion: Consumer Equilibrium

Suppose that as a consumer you have $34 per month to spend on munchies—either pizzas, which cost $6 each, or Twinkies, which cost $4 each.

1. Create a set of marginal utility tables for each product, like the ones below. Remember that they must show diminishing marginal utility as more of each product is consumed.
2. Create the corresponding set of total utility tables for each product.
3. Graph the budget constraint with Pizzas on the horizontal axis and Twinkies on the vertical axis. What are the intercepts? Can you express the budget constraint as an algebraic equation for a line?

# of Pizzas	1	2	3	4	5	6	7
Marginal Utility							
Total Utility							

# of Twinkies	1	2	3	4	5	6	7
Marginal Utility							
Total Utility							

Should you purchase a Twinkie first or a Pizza first to get the "biggest bang for the buck"? How can you tell? What should you purchase second, third, etc. until you exhaust your budget?

Confirm that the combination of Twinkies and Pizzas you end up with will maximize your total utility by computing the total utility from other points on the budget line and comparing them to what you chose.

PART IX

MODULE: PRODUCTION

145. Why It Matters: Production

Why analyze the relationship between inputs used in production, and the resulting outputs and costs?

We are changing gears in this module, which is the first of five dealing with the theory of the firm. The theory of the firm seeks to understand what motivates firms to make the operating decisions that they do. This module, focusing on production and costs (which means the cost of producing a given quantity of output), provides the necessary background information to do that.

This module has a lot of detail, tables and graphs, but there are really two main key takeaways:

1. There are several different ways to look at production and costs, just like there are several different ways to look at a student's learning in a course. Students can be asked to write an essay or do a class presentation on a topic. These are two ways of trying to assess the same thing, student learning.
2. There is an inverse relationship between production and costs. The harder it is to produce something, for example, the more labor it takes, the higher the cost of producing it, and vice versa.

This relationship is illustrated in the following video:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=171#oembed-1

The graphs you will encounter in this module are important because they'll help you keep the different concepts, like production versus costs or different types of cost, straight. Keep track of what's on the horizontal and vertical axis and you'll know what's going on. If employment (or L) is on the horizontal axis, you're looking at some kind of a product curve. If output (or Q) is on the horizontal axis, you're looking at some kind of cost curve.

"Bread Collage". Amazon's Front Door by Robert Scoble, CC-BY. Tesla Autobots by Steve Jurvetson, CC-BY. Road Building by Elvert Barnes, CC-BY-SA. Baking Bread by Christian Guthier, CC-BY. House framing by Ryan, CC-BY-SA.

Students sometimes misunderstand the concept of technology. In an economics context, technology does not mean high tech, or having to do with computers or the internet. Rather, technology means the way, the process by which inputs are used to produce outputs. A different process means a different technology.

Here are some questions to think about: How does an improvement in technology affect production and costs? How, for example, would it affect the graphs? We will answer this at the end of the module, once you have the tools.

One more thing: Do not confuse cost with price. This module is about firms. From the firm's perspective, cost is what they pay for the inputs necessary to produce the product. Price is what the firm receives for selling the product. Thus, cost is a negative and price is positive; they are not the same thing.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=171

LEARNING OUTCOMES

- Define the term "production" and explain what a production function is
- Define and differentiate between marginal, average, and total product; compute and graph marginal, average, and total product
- Differentiate between Explicit and Implicit Costs, Accounting and Economic Profit
- Differentiate between short-run and long-run costs

- Define and explain long-run costs

146. Outcome: Defining Production

What you'll learn to do: define the term "production" and explain what a production function is

In this section, we will learn more about the behavior of firms and how they make production decisions.

The specific things you'll learn to do in this section include:

- Define the term "production inputs," and differentiate between labor, land, capital, entrepreneurship, technology

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Cost and Industry Structure
- Reading: Factors of Production
- Self Check: Defining Production

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

147. Reading: Cost and Industry Structure

Introduction to Cost and Industry Structure

This module is the first of several modules that explore the ***theory of the firm***. This theory explains that firms behave in much the same way as consumers behave. What does that mean? Let's define what is meant by the firm. A ***firm*** (or business) combines inputs of labor, capital, land, and raw or finished component materials to produce outputs. If the firm is successful, the outputs are more valuable than the inputs. This activity of ***production*** goes beyond manufacturing (i.e., making things). It includes any process or service that creates value, including transportation, distribution, wholesale and retail sales. Production involves a number of important decisions that define the behavior of firms. These decisions include, but are not limited to:

- What product or products should the firm produce?
- How should the products be produced (i.e., what production process should be used)?
- How much output should the firm produce?
- What price should the firm charge for its products?
- How much labor should the firm employ?

The answers to these questions depend on the production and cost conditions facing each firm. The answers also depend on the structure of the market for the product(s) in question. Market structure is a multidimensional concept that involves how competitive the industry is. It is defined by questions such as these:

- How much market power does each firm in the industry possess?
- How similar is each firm's product to the products of other firms in the industry?
- How difficult is it for new firms to enter the industry?
- Do firms compete on the basis of price, advertising, or other product differences?

Figure 7.2 illustrates the range of different market structures, which we will explore in Perfect Competition, Monopoly, and Monopolistic Competition and Oligopoly.

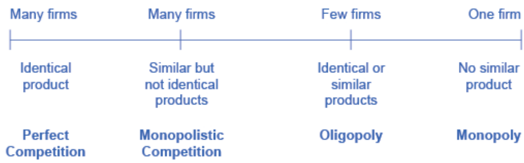


Figure 7.2. *The Spectrum of Competition. Firms face different competitive situations. At one extreme—perfect competition—many firms are all trying to sell identical products. At the other extreme—monopoly—only one firm is selling the product, and this firm faces no competition. Monopolistic competition and oligopoly fall between the extremes of perfect competition and monopoly. Monopolistic competition is a situation with many firms selling similar, but not identical, products. Oligopoly is a situation with few firms that sell identical or similar products.*

148. Reading: Factors of Production

Factors of Production

Choices concerning what goods and services to produce are choices about an economy's use of its factors of production, the resources available to it for the production of goods and services. The value, or satisfaction, that people derive from the goods and services they consume and the activities they pursue is called utility. Ultimately, then, an economy's factors of production create utility; they serve the interests of people.

The factors of production in an economy are its labor, capital, and natural resources. Labor is the human effort that can be applied to the production of goods and services. People who are employed—or are available to be—are considered part of the labor available to the economy. Capital is a factor of production that has been produced for use in the production of other goods and services. Office buildings, machinery, and tools are examples of capital. Natural resources are the resources of nature that can be used for the production of goods and services.

The three basic building blocks of labor, capital, and natural resources may be used in different ways to produce different goods and services, but they still lie at the core of production. We will then look at the roles played by technology and entrepreneurs in putting these factors of production to work. As economists began to grapple with the problems of scarcity, choice, and opportunity cost more than two centuries ago, they focused on these concepts, just as they are likely to do two centuries hence.

Labor

Labor is the human effort that can be applied to production. People who work to repair tires, pilot airplanes, teach children, or enforce laws are all part of the economy's labor. People who would like to work but have not found employment—who are unemployed—are also considered part of the labor available to the economy.

In some contexts, it is useful to distinguish two forms of labor. The first is the human equivalent of a natural resource. It is the natural ability an untrained, uneducated person brings to a particular production process. But most workers bring far more. Skills a worker has as a result of education, training, or experience that can be used in production are called human capital. Students are acquiring human capital. Workers who are gaining skills through experience or through training are acquiring human capital.

The amount of labor available to an economy can be increased in two ways. One is to increase the total quantity of labor, either by increasing the number of people available to work or by increasing the average number of hours of work per time period. The other is to increase the amount of human capital possessed by workers.

Capital

Long ago, when the first human beings walked the earth, they produced food by picking leaves or fruit off a plant or by catching an animal and eating it. We know that very early on, however, they began shaping stones into tools, apparently for use in butchering animals. Those tools were the first capital because they were produced for use in producing other goods—food and clothing.

Modern versions of the first stone tools include saws, meat cleavers, hooks, and grinders; all are used in butchering animals. Tools such as hammers, screwdrivers, and wrenches are also capital.

Transportation equipment, such as cars and trucks, is capital. Facilities such as roads, bridges, ports, and airports are capital. Buildings, too, are capital; they help us to produce goods and services.

Capital does not consist solely of physical objects. The score for a new symphony is capital because it will be used to produce concerts. Computer software used by business firms or government agencies to produce goods and services is capital. Capital may thus include physical goods and intellectual discoveries. Any resource is capital if it satisfies two criteria:

1. The resource must have been produced.
2. The resource can be used to produce other goods and services.

One thing that is not considered capital is money. A firm cannot use money directly to produce other goods, so money does not satisfy the second criterion for capital. Firms can, however, use money to acquire capital. Money is a form of financial capital. Financial capital includes money and other "paper" assets (such as stocks and bonds) that represent claims on future payments. These financial assets are not capital, but they can be used directly or indirectly to purchase factors of production or goods and services.

Natural Resources

There are two essential characteristics of natural resources. The first is that they are found in nature—that no human effort has been used to make or alter them. The second is that they can be used for the production of goods and services. That requires knowledge; we must know how to use the things we find in nature before they become resources.

Consider oil. Oil in the ground is a natural resource because it is found (not manufactured) and can be used to produce goods and services. However, 250 years ago oil was a nuisance, not a natural resource. Pennsylvania farmers in the eighteenth century who found oil oozing up through their soil were dismayed, not delighted. No one knew what could be done with the oil. It was not until the mid-nineteenth century that a method was found for refining oil into kerosene that could be used to generate energy, transforming oil into a natural resource. Oil is now used to make all sorts of things, including clothing, drugs, gasoline, and plastic. It became a natural resource because people discovered and implemented a way to use it.

Defining something as a natural resource only if it can be used to produce goods and services does not mean that a tree has value only for its wood or that a mountain has value only for its minerals. If people gain utility from the existence of a beautiful wilderness area, then that wilderness provides a service. The wilderness is thus a natural resource.

The natural resources available to us can be expanded in three ways. One is the discovery of new natural resources, such as the discovery of a deposit of ore containing titanium. The second is the discovery of new uses for resources, as happened when new techniques allowed oil to be put to productive use or sand to be used in manufacturing computer chips. The third is the discovery of new ways to extract natural resources in order to use them. New methods of discovering and mapping oil deposits have increased the world's supply of this important natural resource.

Technology and the Entrepreneur

Goods and services are produced using the factors of production available to the economy. Two things play a crucial role in putting these factors of production to work. The first is technology, the knowledge that can be applied to the production of goods and services. The second is an individual who plays a key role in a market economy: the entrepreneur. An entrepreneur is a person who, operating within the context of a market economy, seeks to earn profits by finding new ways to organize factors of production. In non-market economies the role of the entrepreneur is played by bureaucrats and other decision makers who respond to incentives other than profit to guide their choices about resource allocation decisions.

The interplay of entrepreneurs and technology affects all our lives. Entrepreneurs put new technologies to work every day, changing the way factors of production are used. Farmers and factory workers, engineers and electricians, technicians and teachers all work differently than they did just a few years ago, using new technologies introduced by entrepreneurs. The music you enjoy, the books you read, the athletic equipment with which you play are produced differently than they were five years ago. The text you are reading was written and manufactured using technologies that did not exist ten years ago. We can dispute whether all the changes have made our lives better. What we cannot dispute is that they have made our lives different.

Self Check: Defining Production

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/640

149. Outcome: Marginal, Average, and Total Product

What you'll learn to do: define and differentiate between marginal, average, and total product; compute and graph marginal, average, and total product

In this section, you'll learn more about the types of decisions that firms face in the short-run. A firm must consider their marginal, average, and total products to help determine how much they could and should produce.

The specific things you'll learn to do in this section include:

- Explain diminishing marginal product and diminishing marginal returns

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Production Choices and Costs
- Self Check: Marginal, Average, and Total Product

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

150. Reading: Production Choices and Costs

Production Choices and Costs

Our analysis of production and cost begins with a period economists call the short run. The short run in this microeconomic context is a planning period over which the managers of a firm must consider one or more of their factors of production as fixed in quantity. For example, a restaurant may regard its building as a fixed factor over a period of at least the next year. It would take at least that much time to find a new building or to expand or reduce the size of its present facility. Decisions concerning the operation of the restaurant during the next year must assume the building will remain unchanged. Other factors of production could be changed during the year, but the size of the building must be regarded as a constant.

When the quantity of a factor of production cannot be changed during a particular period, it is called a fixed factor of production. For the restaurant, its building is a fixed factor of production for at least a year. A factor of production whose quantity can be changed during a particular period is called a variable factor of production; factors such as labor and food are examples.

While the managers of the restaurant are making choices concerning its operation over the next year, they are also planning for longer periods. Over those periods, managers may contemplate alternatives such as modifying the building, building a new facility, or selling the building and leaving the restaurant business. The planning period over which a firm can consider *all* factors of production as variable is called the long run.

At any one time, a firm will be making both short-run and long-run choices. The managers may be planning what to do for the next few weeks and for the next few years. Their decisions over the next few weeks are likely to be short-run choices. Decisions that will affect operations over the next few years may be long-run choices, in which managers can consider changing every aspect of their operations. Our analysis in this section focuses on the short run. We examine long-run choices later in this module.

The Short-Run Production Function

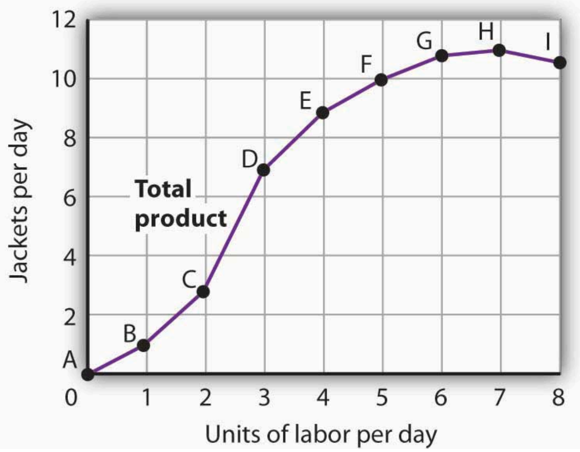
A firm uses factors of production to produce a product. The relationship between factors of production and the output of a firm is called a production function. Our first task is to explore the nature of the production function.

Consider a hypothetical firm, Acme Clothing, a shop that produces jackets. Suppose that Acme has a lease on its building and equipment. During the period of the lease, Acme's capital is its fixed factor of production. Acme's variable factors of production include things such as labor, cloth, and electricity. In the analysis that follows, we shall simplify by assuming that labor is Acme's *only* variable factor of production.

Total, Marginal and Average Products

Figure 8.1 "Acme Clothing's Total Product Curve" shows the number of jackets Acme can obtain with varying amounts of labor (in this case, tailors) and its given level of capital. A total product curve shows the quantities of output that can be obtained from different amounts of a variable factor of production, assuming other factors of production are fixed.

Point on graph	A	B	C	D	E	F	G	H	I
Units of labor per day	0	1	2	3	4	5	6	7	8
Jackets per day	0.0	1.0	3.0	7.0	9.0	10.0	10.7	11.0	10.5



Notice what happens to the slope of the total product curve in Figure 8.1 "Acme Clothing's Total Product Curve." Between 0 and 3 units of labor per day, the curve becomes steeper. Between 3 and 7 workers, the curve continues to slope upward, but its slope diminishes. Beyond the seventh tailor, production begins to decline and the curve slopes downward.

We measure the slope of any curve as the vertical change between two points divided by the horizontal change between the same two points. The slope of the total product curve for labor equals the change in output (ΔQ) divided by the change in units of labor (ΔL):

The slope of a total product curve for any variable factor is a measure of the change in output associated with a change in the amount of the variable factor, with the quantities of all other factors held constant. The amount by which output rises with an additional unit of a variable factor is the marginal product of the variable factor. Mathematically, marginal product is the ratio of the change in output to the change in the amount of a variable factor. The marginal product of labor (MP_L), for example, is the amount by which output rises with an additional unit of labor. It is thus the ratio of the change in output to the change in the quantity of labor

($\Delta Q/\Delta L$), all other things unchanged. It is measured as the slope of the total product curve for labor.

Equation 8.1

$$MP_L = \frac{\Delta Q}{\Delta L}$$

In addition we can define the average product of a variable factor. It is the output per unit of variable factor. The average product of labor (AP_L), for example, is the ratio of output to the number of units of labor (Q/L).

Equation 8.2

$$AP_L = \frac{Q}{L}$$

The concept of average product is often used for comparing productivity levels over time or in comparing productivity levels among nations. When you read in the newspaper that productivity is rising or falling, or that productivity in the United States is nine times greater than productivity in China, the report is probably referring to some measure of the average product of labor.

The total product curve in Panel (a) of Figure 8.2 "From Total Product to the Average and Marginal Product of Labor" is repeated from Figure 8.1 "Acme Clothing's Total Product Curve". Panel (b) shows the marginal product and average product curves. Notice that marginal product is the slope of the total product curve, and that marginal product rises as the slope of the total product curve increases, falls as the slope of the total product curve declines, reaches zero when the total product curve achieves its maximum value, and becomes negative as the total product curve slopes downward. As in other parts of this text, marginal values are plotted at the midpoint of each interval. The marginal product of the fifth unit of labor, for example, is plotted between 4 and 5 units of labor. Also notice that the marginal product curve intersects the average product curve at the maximum point on the average product curve. When marginal product is above average product, average product is rising. When marginal product is below average product, average product is falling.

Units of labor per day	0	1	2	3	4	5	6	7	8
Jackets per day	0	1.0	3.0	7.0	9.0	10.0	10.7	11.0	10.5
Marginal product		1.0	2.0	4.0	2.0	1.0	0.7	0.3	−0.5
Average product	–	1.0	1.5	2.33	2.25	2.0	1.78	1.57	1.31

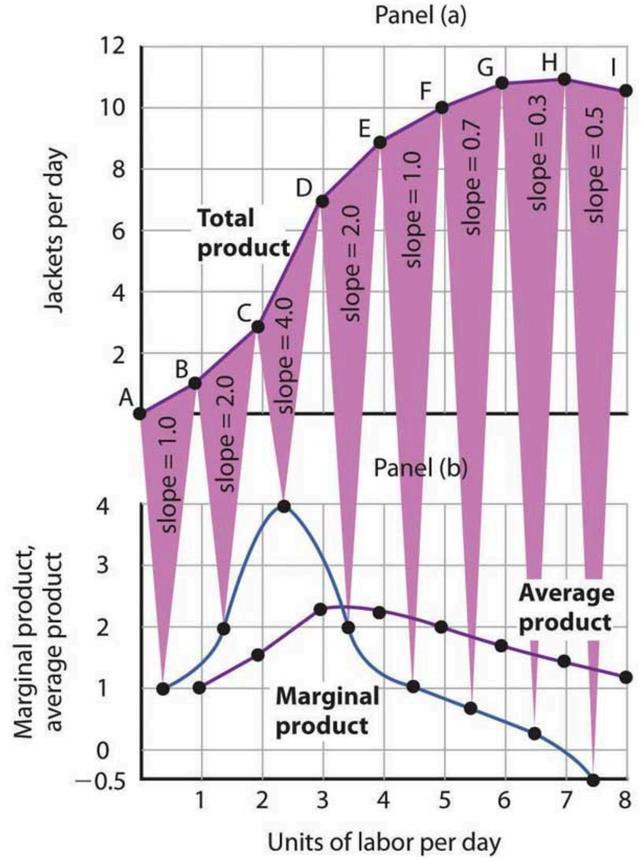


Figure 8.2 From Total Product to the Average and Marginal Product of Labor. *The first two rows of the table give the values for quantities of labor and total product from Figure 8.1 "Acme Clothing's Total Product Curve." Marginal product, given in the third row, is the change in output resulting from a one-unit increase in labor. Average product, given in the fourth row, is output per unit of labor. Panel (a) shows the total product curve. The slope of the total product curve is marginal product, which is plotted in Panel (b). Values for marginal product are plotted at the midpoints of the intervals. Average product rises and falls. Where marginal product is above average product, average product rises. Where marginal product is below average product,*

average product falls. The marginal product curve intersects the average product curve at the maximum point on the average product curve.

As a student you can use your own experience to understand the relationship between marginal and average values. Your grade point average (GPA) represents the average grade you have earned in all your course work so far. When you take an additional course, your grade in that course represents the marginal grade. What happens to your GPA when you get a grade that is higher than your previous average? It rises. What happens to your GPA when you get a grade that is lower than your previous average? It falls. If your GPA is a 3.0 and you earn one more B, your marginal grade equals your GPA and your GPA remains unchanged.

The relationship between average product and marginal product is similar. However, unlike your course grades, which may go up and down willy-nilly, marginal product always rises and then falls, for reasons we will explore shortly. As soon as marginal product falls below average product, the average product curve slopes downward. While marginal product is above average product, whether marginal product is increasing or decreasing, the average product curve slopes upward.

As we have learned, maximizing behavior requires focusing on making decisions at the margin. For this reason, we turn our attention now toward increasing our understanding of marginal product.

Increasing, Diminishing and Negative Marginal Returns

Adding the first worker increases Acme's output from 0 to 1 jacket per day. The second tailor adds 2 jackets to total output; the third adds 4. The marginal product goes up because when there are more workers, each one can specialize to a degree. One worker might cut the cloth, another might sew the seams, and another might sew the buttonholes. Their increasing marginal products are reflected by the increasing slope of the total product curve over the first 3 units of labor and by the upward slope of the marginal product curve over the same range. The range over which marginal products are increasing is called the range of increasing marginal returns. Increasing marginal returns exist in the context of a total product curve for labor, so we are holding the quantities of other factors constant. Increasing marginal returns may occur for any variable factor.

The fourth worker adds less to total output than the third; the marginal product of the fourth worker is 2 jackets. The data in Figure 8.2 "From Total Product to the Average and Marginal Product of Labor" show that marginal product continues to decline after the fourth worker as more and more workers are hired. The additional workers allow even greater opportunities for specialization, but because they are operating with a fixed amount of capital, each new worker adds less to total output. The fifth tailor adds only a single jacket to total output. When each additional unit of a variable factor adds less to total output, the firm is experiencing diminishing marginal returns. Over the range of diminishing marginal returns, the marginal product of the variable factor is positive but falling. Once again, we assume that the quantities of all other factors of production are fixed. Diminishing marginal returns may occur for any variable factor. Panel (b) shows that Acme experiences diminishing marginal returns between the third and seventh workers, or between 7 and 11 jackets per day.

After the seventh unit of labor, Acme's fixed plant becomes so crowded that adding another worker actually reduces output. When additional units of a variable factor reduce total output, given constant quantities of all other factors, the company experiences negative marginal returns. Now the total product curve is downward sloping, and the marginal product curve falls below zero. Figure 8.3 "Increasing Marginal Returns, Diminishing Marginal Returns, and Negative Marginal Returns" shows the ranges of increasing, diminishing, and negative marginal returns. Clearly, a firm will never intentionally add so much of a variable factor of production that it enters a range of negative marginal returns.

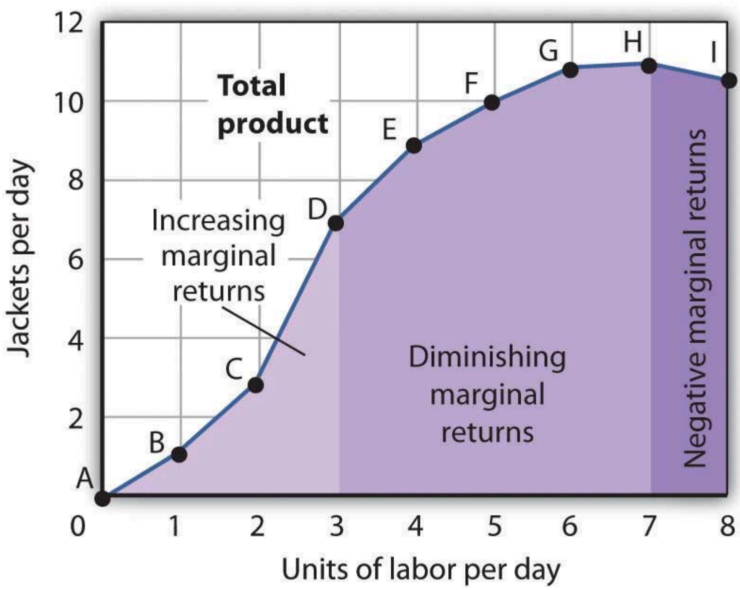


Figure 8.3 Increasing Marginal Returns, Diminishing Marginal Returns, and Negative Marginal Returns. *This graph shows Acme's total product curve from Figure 8.1 "Acme Clothing's Total Product Curve" with the ranges of increasing marginal returns, diminishing marginal returns, and negative marginal returns marked. Acme experiences increasing marginal returns between 0 and 3 units of labor per day, diminishing marginal returns between 3 and 7 units of labor per day, and negative marginal returns beyond the 7th unit of labor.*

The idea that the marginal product of a variable factor declines over some range is important enough, and general enough, that economists state it as a law. The law of diminishing marginal returns holds that the marginal product of any variable factor of production will eventually decline, assuming the quantities of other factors of production are unchanged.

Heads Up!

It is easy to confuse the concept of diminishing marginal returns with the idea of negative marginal returns. To say a firm is experiencing diminishing marginal returns is not to say its output is falling. Diminishing marginal returns mean that the marginal product of a variable factor is declining. Output is still increasing as the variable factor is increased, but it is increasing by smaller and smaller amounts. As we saw in Figure 8.2 “From Total Product to the Average and Marginal Product of Labor” and Figure 8.3 “Increasing Marginal Returns, Diminishing Marginal Returns, and Negative Marginal Returns,” the range of diminishing marginal returns was between the third and seventh workers; over this range of workers, output rose from 7 to 11 jackets. Negative marginal returns started after the seventh worker.

To see the logic of the law of diminishing marginal returns, imagine a case in which it does not hold. Say that you have a small plot of land for a vegetable garden, 10 feet by 10 feet in size. The plot itself is a fixed factor in the production of vegetables. Suppose you are able to hold constant all other factors—water, sunshine, temperature, fertilizer, and seed—and vary the amount of labor devoted to the garden. How much food could the garden produce? Suppose the marginal product of labor kept increasing or was constant. Then you could grow an *unlimited* quantity of food on your small plot—enough to feed the entire world! You could add an unlimited number of workers to your plot and still increase output at a constant or increasing rate. If you did not get enough output with, say, 500 workers, you could use 5 million; the five-millionth worker would add at least as much to total output as the first.

If diminishing marginal returns to labor did not occur, the total product curve would slope upward at a constant or increasing rate.

The shape of the total product curve and the shape of the resulting marginal product curve drawn in Figure 8.2 “From Total Product to the Average and Marginal Product of Labor” are typical of *any* firm for the short run. Given its fixed factors of production, increasing the use of a variable factor will generate increasing marginal returns at first; the total product curve for the variable factor becomes steeper and the marginal product rises. The opportunity to gain from increased specialization in the use of the variable factor accounts for this range of increasing marginal returns. Eventually, though, diminishing returns will set in. The total product curve will become flatter, and the marginal product curve will fall.

Self Check: Marginal, Average, and Total Product

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You’ll have more success on the Self Check if you’ve completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/762

151. Outcome: Explicit and Implicit Costs

What you'll learn to do: differentiate between explicit and implicit costs, accounting, and economic profit.

If you are interested in someday starting your own restaurant, you'll definitely need to prepare for explicit costs—the obvious costs of paying rent, buying ingredients, paying your staff, etc. What will be less obvious and harder to measure, however, will be the implicit costs—the sleepless nights you spend working on the design, the salary you don't make right off the bat, and the ingredients you experiment with to create your special sauce. In this section, you'll learn more about these types of costs.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Explicit and Implicit Costs
- Self Check: Explicit and Implicit Costs

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

152. Reading: Explicit and Implicit Costs

Explicit and Implicit Costs, and Accounting and Economic Profit

Private enterprise, the ownership of businesses by private individuals, is a hallmark of the U.S. economy. When people think of businesses, often giants like Wal-Mart, Microsoft, or General Motors come to mind. But firms come in all sizes, as shown in Table 2.1. The vast majority of American firms have fewer than 20 employees. As of 2010, the U.S. Census Bureau counted 5.7 million firms with employees in the U.S. economy. Slightly less than half of all the workers in private firms are at the 17,000 large firms, meaning they employ more than 500 workers. Another 35% of workers in the U.S. economy are at firms with fewer than 100 workers. These small-scale businesses include everything from dentists and lawyers to businesses that mow lawns or clean houses. Indeed, Table 2.1 does not include a separate category for the millions of small "non-employer" businesses where a single owner or a few partners are not officially paid wages or a salary, but simply receive whatever they can earn.

Table 2.1 Range in Size of U.S. Firms (Source: U.S. Census, 2010 www.census.gov)

Number of Employees	Firms (% of total firms)	Number of Paid Employees (% of total employment)
Total	5,734,538	112.0 million
0–9	4,543,315 (79.2%)	12.3 million (11.0%)
10–19	617,089 (10.8%)	8.3 million (7.4%)
20–99	475,125 (8.3%)	18.6 million (16.6%)
100–499	81,773 (1.4%)	15.9 million (14.2%)
500 or more	17,236 (0.30%)	50.9 million (49.8%)

Each of these businesses, regardless of size or complexity, tries to earn a profit:

Profit = Total Revenue − Total Cost

Total revenue is the income brought into the firm from selling its products. It is calculated by multiplying the price of the product times the quantity of output sold:

Total Revenue = Price x Quantity

We will see in the following modules that revenue is a function of the demand for the firm's products.

We can distinguish between two types of cost: explicit and implicit. Explicit costs are out-of-pocket costs, that is, payments that are actually made. Wages that a firm pays its employees or rent that a firm pays for its office are explicit costs. Implicit costs are more subtle, but just as important. They represent the opportunity cost of using resources already owned by the firm. Often for small businesses, they are resources contributed by the owners; for example, working in the business while not getting a formal salary, or using the ground floor of a home as a retail store. Implicit costs also allow for depreciation of goods, materials, and equipment that are necessary for a company to operate.

These two definitions of cost are important for distinguishing between two conceptions of profit, accounting profit and economic profit. Accounting profit is a cash concept. It means total revenue minus explicit costs—the difference between dollars brought in and dollars paid out. Economic profit is total revenue minus total cost, including both explicit and implicit costs. The difference is important because even though a business pays income taxes based on its accounting profit, whether or not it is economically successful depends on its economic profit.

CALCULATING IMPLICIT COSTS

Consider the following example. Fred currently works for a corporate law firm. He is considering opening his own legal practice, where he expects to earn $200,000 per year once he gets established. To run his own firm, he would need an office and a law clerk. He has found the perfect office, which rents for $50,000 per year. A law clerk could be hired for $35,000 per year. If these figures are accurate, would Fred's legal practice be profitable?

Step 1. First you have to calculate the costs. You can take what you know about explicit costs and total them:

Office rental	$50,000
Law clerk's salary	+ $35,000
Total explicit costs	$85,000

Step 2. Subtracting the explicit costs from the revenue gives you the accounting profit.

Revenues	$200,000
Explicit costs	– $85,000
Accounting profit	$115,000

But these calculations consider only the explicit costs. To open his own practice, Fred would have to quit his current job, where he is earning an annual salary of $125,000. This would be an implicit cost of opening his own firm.

Step 3. You need to subtract both the explicit and implicit costs to determine the true economic profit. The equation is:

$$\text{Economic Profit} = \text{Total Revenues} - \text{Explicit Costs} - \text{Implicit Costs}$$

Now let's plug in Fred's figures to the true economic profit equation:

$$\begin{aligned}\text{Economic Profit} &= \$200{,}000 - \$85{,}000 - \$125{,}000 \\ &= -\$10{,}000 \text{ per year}\end{aligned}$$

Fred would be losing $10,000 per year. That does not mean he would not want to open his own business, but it does mean he would be earning $10,000 less than if he worked for the corporate firm.

Implicit costs can include other things as well. Maybe Fred values his leisure time, and starting his own firm would require him to put in more hours than at the corporate firm. In this case, the lost leisure would also be an implicit cost that would subtract from economic profits.

Now that we have an idea about the different types of costs, let's look at cost structures. A firm's cost structure in the long run may be different from that in the short run. We turn to that distinction in the next section.

Self Check: Explicit and Implicit Costs

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/644

153. Outcome: Marginal, Average, and Total Cost

What you'll learn to do: define and differentiate between marginal, average, and total cost; compute and graph marginal, average, and total cost

In this section, you'll see how firms look at marginal, average, and total costs curves to determine whether or not it is making a profit, or at least headed in that direction.

The specific things you'll learn to do in this section include:

- Differentiate between variable and fixed costs

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Fixed and Variable Costs
- Reading: The Structure of Costs in the Short Run
- Self Check: Marginal, Average, and Total Cost

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

154. Reading: Fixed and Variable Costs

Fixed and Variable Costs

Fixed costs are expenditures that do not change regardless of the level of production, at least not in the short term. Whether you produce a lot or a little, the fixed costs are the same. One example is the rent on a factory or a retail space. Once you sign the lease, the rent is the same regardless of how much you produce, at least until the lease runs out. Fixed costs can take many other forms: for example, the cost of machinery or equipment to produce the product, research and development costs to develop new products, even an expense like advertising to popularize a brand name. The level of fixed costs varies according to the specific line of business: for instance, manufacturing computer chips requires an expensive factory, but a local moving and hauling business can get by with almost no fixed costs at all if it rents trucks by the day when needed. *Variable costs*, on the other hand, are incurred in the act of producing—the more you produce, the greater the variable cost. Labor is treated as a variable cost, since producing a greater quantity of a good or service typically requires more workers or more work hours. Variable costs would also include raw materials.

As a concrete example of fixed and variable costs, consider the barber shop called "The Clip Joint" shown in Figure 7.3. The data for output and costs are shown in Table 7.2. The fixed costs of operating the barber shop, including the space and equipment, are $160 per day. The variable costs are the costs of hiring barbers, which in our example is $80 per barber each day. The first two columns of the table show the quantity of haircuts the barbershop can produce as it hires additional barbers. The third column shows the fixed costs, which do not change regardless of the level of production. The fourth column shows the variable costs at each level of output. These are calculated by taking the amount of labor hired and multiplying by the wage. For example, two barbers cost: 2 × $80 = $160. Adding together the fixed costs in the third column and the variable costs in the fourth column produces the total costs in the fifth column. So, for example, with two barbers the total cost is: $160 + $160 = $320.

Table 7.2. Output and Total Costs

Labor	**Quantity**	**Fixed Cost**	**Variable Cost**	**Total Cost**
1	16	$160	$80	$240
2	40	$160	$160	$320
3	60	$160	$240	$400
4	72	$160	$320	$480
5	80	$160	$400	$560
6	84	$160	$480	$640
7	82	$160	$560	$720

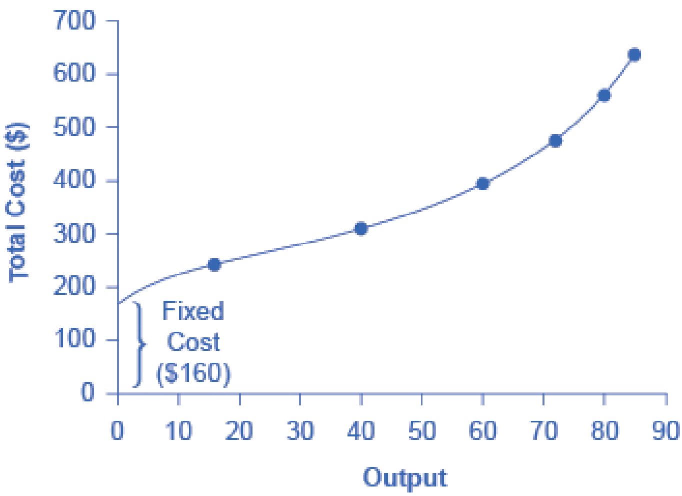


Figure 7.3. How Output Affects Total Costs *At zero production, the fixed costs of $160 are still present. As production increases, variable costs are added to fixed costs, and the total cost is the sum of the two.*

The relationship between the quantity of output being produced and the cost of producing that output is shown graphically in the figure. The fixed costs are always shown as the vertical intercept of the total cost curve; that is, they are the costs incurred when output is zero so there are no variable costs. You can see from the graph that once production starts, total costs and variable costs rise. While variable costs may initially increase at a decreasing rate, at some point they begin increasing at an increasing rate. This is caused by diminishing marginal returns, discussed in the module on Choice in a World of Scarcity, which is easiest to see with an example. As the number of barbers increases from zero to one in the table, output increases from 0 to 16 for a marginal gain of 16; as the number rises from one to two barbers, output increases from 16 to 40, a marginal gain of 24. From that point on, though, the marginal gain in output diminishes as each additional barber is added.

For example, as the number of barbers rises from two to three, the marginal output gain is only 20; and as the number rises from three to four, the marginal gain is only 12. To understand the reason behind this pattern, consider that a one-man barber shop is a very busy operation. The single barber needs to do everything: say hello to people entering, answer the phone, cut hair, sweep up, and run the cash register. A second barber reduces the level of disruption from jumping back and forth between these tasks, and allows a greater division of labor and specialization. The result can be greater increasing marginal returns. However, as other barbers are added, the advantage of each additional barber is less, since the specialization of labor can only go so far. The addition of a sixth or seventh or eighth barber just to greet people at the door will have less impact than the second one did. This is the pattern of diminishing marginal returns. At some point, you may even see negative returns as the additional barbers begin bumping elbows and getting in each other's way. In this case, the addition of still more barbers would actually cause output to decrease, as shown in the last row of Table 7.2. As a result, the total costs of production will begin to rise more rapidly as output increases.

This pattern of *diminishing marginal returns* is common in production. As another example, consider the problem of irrigating a crop on a farmer's field. The plot of land is the fixed factor of production, while the water that can be added to the land is the key variable cost. As the farmer adds water to the land, output increases. But adding more and more water brings smaller and smaller increases in output, until at some point the water floods the field and actually reduces output. Diminishing marginal returns occur because, at a given level of fixed costs, each additional input contributes less and less to overall production.

155. Reading: The Structure of Costs in the Short Run

The Structure of Costs in the Short Run

The cost of producing a firm's output depends on how much *labor* and *physical capital* the firm uses. A list of the costs involved in producing cars will look very different from the costs involved in producing computer software or haircuts or fast-food meals. However, the cost structure of all firms can be broken down into some common underlying patterns. When a firm looks at its *total costs* of production in the short run, a useful starting point is to divide total costs into two categories: fixed costs that cannot be changed in the short run and variable costs that can be changed.

The breakdown of total costs into fixed and variable costs can provide a basis for other insights as well. The first five columns of Table 7.3 duplicate the previous table, but the last three columns show average total costs, average variable costs, and marginal costs. These new measures analyze costs on a per-unit (rather than a total) basis and are reflected in the curves shown in Figure 7.4.

Table 7.3. Output and Total Costs

Labor	Quantity	Fixed Cost	Variable Cost	Total Cost
1	16	$160	$80	$240
2	40	$160	$160	$320
3	60	$160	$240	$400
4	72	$160	$320	$480
5	80	$160	$400	$560
6	84	$160	$480	$640
7	82	$160	$560	$720

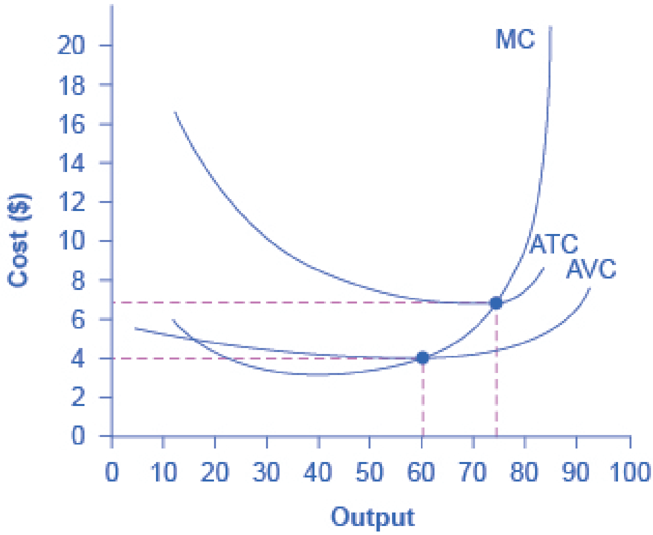


Figure 7.4. *Cost Curves at the Clip Joint. The information on total costs, fixed cost, and variable cost can also be presented on a per-unit basis. Average total cost (ATC) is calculated by dividing total cost by the total quantity produced. The average total cost curve is typically U-shaped. Average variable cost (AVC) is calculated by dividing variable cost by the quantity produced. The average variable cost curve lies below the average total cost curve and is typically U-shaped or upward-sloping. Marginal cost (MC) is calculated by taking the change in total cost between two levels of output and dividing by the change in output. The marginal cost curve is upward-sloping.*

Average total cost is total cost divided by the quantity of output. Since the total cost of producing 40 haircuts is $320, the average total cost for producing each of 40 haircuts is $320/40, or $8 per haircut. Average cost curves are typically U-shaped, as Figure 7.4 shows. Average total cost starts off relatively high, because at low levels of output total costs are dominated by the fixed cost; mathematically, the denominator is so small that average total cost is large. Average total cost then declines, as the fixed costs are spread over an increasing quantity of output. In the average cost calculation, the rise in the numerator of total costs is relatively small compared to the rise in the denominator of quantity produced. But as output expands still further, the average cost begins to rise. At the right side of the average cost curve, total costs begin rising more rapidly as diminishing returns kick in.

Average variable cost obtained when variable cost is divided by quantity of output. For example, the variable cost of producing 80 haircuts is $400, so the average variable cost is $400/80, or $5 per haircut. Note that at any level of output, the average variable cost curve will always lie below the curve for average total cost, as shown in Figure 7.4. The reason is that average total cost includes average variable cost and average fixed cost. Thus, for Q = 80 haircuts, the average total cost is $8 per haircut, while the average variable cost is $5 per haircut. However, as output grows, fixed costs become relatively less important (since they do not rise with output), so average variable cost sneaks closer to average cost. Average total and variable costs measure the average costs of producing some quantity of output. Marginal cost is somewhat different.

Marginal cost is the additional cost of producing one more unit of output. So it is not the cost per unit of *all* units being produced, but only the next one (or next few). Marginal cost can be calculated by taking the change in total cost and dividing it by the change in quantity. For example, as quantity produced increases from 40 to 60 haircuts, total costs rise by 400 – 320, or 80. Thus, the marginal cost for each of those marginal 20 units will be 80/20, or $4 per haircut. The marginal cost curve is generally upward-sloping, because diminishing marginal returns implies that additional units are more costly to produce. A small range of increasing marginal returns can be seen in the figure as a dip in the marginal cost curve before it starts rising. There is a point at which marginal and average costs meet, as explained below.

Where do marginal and average costs meet?

The marginal cost line intersects the average cost line exactly at the bottom of the average cost curve—which occurs at a quantity of 72 and cost of \$6.60 in Figure 7.4. The reason why the intersection occurs at this point is built into the economic meaning of marginal and average costs. If the marginal cost of production is below the average cost for producing previous units, as it is for the points to the left of where MC crosses ATC, then producing one more additional unit will reduce average costs overall—and the ATC curve will be downward-sloping in this zone. Conversely, if the marginal cost of production for producing an additional unit is above the average cost for producing the earlier units, as it is for points to the right of where MC crosses ATC, then producing a marginal unit will increase average costs overall—and the ATC curve must be upward-sloping in this zone. The point of transition, between where MC is pulling ATC down and where it is pulling it up, must occur at the minimum point of the ATC curve.

This idea of the marginal cost "pulling down" the average cost or "pulling up" the average cost may sound abstract, but think about it in terms of your own grades. If the score on the most recent quiz you take is lower than your average score on previous quizzes, then the marginal quiz pulls down your average. If your score on the most recent quiz is higher than the average on previous quizzes, the marginal quiz pulls up your average. In this same way, low marginal costs of production first pull down average costs and then higher marginal costs pull them up.

The numerical calculations behind average cost, average variable cost, and marginal cost will change from firm to firm. However, the general patterns of these curves, and the relationships and economic intuition behind them, will not change.

Self Check: Marginal, Average, and Total Cost

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/763

156. Outcome: Sunk Costs

What you'll learn to do: identify sunk costs

It's obvious that a firm's total revenue must exceed total costs if it wants to make a profit. But in this section, you'll see why it's helpful for firms to break down their costs to examine various types of costs—fixed cost, marginal cost, average total cost, average variable costs, and sunk costs to get a better picture of how they can become more profitable.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Sunk Costs and Alternative Measures of Costs
- Self Check: Sunk Costs

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

157. Reading: Sunk Costs and Alternative Measures of Cost

LESSONS FROM ALTERNATIVE MEASURES OF COSTS

Breaking down total costs into fixed cost, marginal cost, average total cost, and average variable cost is useful because each statistic offers its own insights for the firm. Whatever the firm's quantity of production, total revenue must exceed total costs if it is to earn a profit.

As explored in the section Choice in a World of Scarcity of the Introduction to Economics and Scarcity module, fixed costs are often *sunk costs* that cannot be recouped. In thinking about what to do next, sunk costs should typically be ignored, since this spending has already been made and cannot be changed. However, variable costs can be changed, so they convey information about the firm's ability to cut costs in the present and the extent to which costs will increase if production rises.

Why are total cost and average cost not on the same graph?

Total cost, fixed cost, and variable cost each reflect different aspects of the cost of production over the entire quantity of output being produced. These costs are measured in dollars. In contrast, marginal cost, average cost, and average variable cost are costs per unit. In the previous example, they are measured as cost per haircut. Thus, it would not make sense to put all of these numbers on the same graph, since they are measured in different units ($ versus $ per unit of output).

It would be as if the vertical axis measured two different things. In addition, as a practical matter, if they were on the same graph, the lines for marginal cost, average cost, and average variable cost would appear almost flat against the horizontal axis, compared to the values for total cost, fixed cost, and variable cost. Using the figures from the previous example, the total cost of producing 40 haircuts is $320. But the average cost is $320/40, or $8. If you graphed both total and average cost on the same axes, the average cost would hardly show.

Average cost tells a firm whether it can earn profits given the current price in the market. If we divide profit by the quantity of output produced we get *average profit*, also known as the firm's *profit margin*. Expanding the equation for profit gives:

$$\text{average profit} = \frac{\text{profit}}{\text{quantity produced}}$$

$$\text{average profit} = \frac{\text{total revenue} - \text{total cost}}{\text{quantity produced}}$$

$$\text{average profit} = \frac{\text{total revenue}}{\text{quantity produced}} - \frac{\text{total cost}}{\text{quantity produced}}$$

$$\text{average profit} = \text{average revenue} - \text{average cost}$$

But note that:

$$\text{average revenue} = \frac{\text{price} \times \text{quantity produced}}{\text{quantity produced}}$$

$$\text{average revenue} = \text{price}$$

Thus:

$$\text{average profit} = \text{price} - \text{average cost}$$

This is the firm's *profit margin*. This definition implies that if the market price is above average cost, average profit, and thus total profit, will be positive; if price is below average cost, then profits will be negative. The marginal cost of producing an additional unit can be compared with the marginal revenue gained by selling that additional unit to reveal whether the additional unit is adding to total profit—or not. Thus, marginal cost helps producers understand how profits would be affected by increasing or decreasing production.

A VARIETY OF COST PATTERNS

The pattern of costs varies among industries and even among firms in the same industry. Some businesses have high fixed costs, but low marginal costs. Consider, for example, an Internet company that provides medical advice to customers. Such a company might be paid by consumers directly, or perhaps hospitals or healthcare practices might subscribe on behalf of their patients.

Setting up the website, collecting the information, writing the content, and buying or leasing the computer space to handle the web traffic are all fixed costs that must be undertaken before the site can work. However, when the website is up and running, it can provide a high quantity of service with relatively low variable costs, like the cost of monitoring the system and updating the information.

In this case, the total cost curve might start at a high level, because of the high fixed costs, but then might appear close to flat, up to a large quantity of output, reflecting the low variable costs of operation. If the website is popular, however, a large rise in the number of visitors will overwhelm the website, and increasing output further could require a purchase of additional computer space.

For other firms, fixed costs may be relatively low. For example, consider firms that rake leaves in the fall or shovel snow off sidewalks and driveways in the winter. For fixed costs, such firms may need little more than a car to transport workers to homes of customers and some rakes and shovels. Still other firms may find that diminishing marginal returns set in quite sharply. If a manufacturing plant tried to run 24 hours a day, seven days a week, little time remains for routine maintenance of the equipment, and marginal costs can increase dramatically as the firm struggles to repair and replace overworked equipment.

Every firm can gain insight into its task of earning profits by dividing its total costs into fixed and variable costs, and then using these calculations as a basis for average total cost, average variable cost, and marginal cost. However, making a final decision about the profit-maximizing quantity to produce and the price to charge will require combining these perspectives on cost with an analysis of sales and revenue, which in turn requires looking at the market structure in which the firm finds itself. Before we turn to the analysis of market structure in other modules, we will analyze the firm's cost structure from a long-run perspective.

Self Check: Sunk Costs

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/645

158. Outcome: The Short Run vs. The Long Run

What you'll learn to do: differentiate between short-run and long-run costs

In this section, you will see how time is an important factor to consider in a firm's success.

The specific things you'll learn to do in this section include:

- Interpret the relationship between short run and long run costs.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Short Run and Long Run Average Total Costs
- Self Check: The Short Run vs. The Long Run

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

159. Reading: Short Run vs. Long Run Costs

Short Run vs. Long Run Costs

Our analysis of production and cost begins with a period economists call the short run. The short run in this microeconomic context is a planning period over which the managers of a firm must consider one or more of their factors of production as fixed in quantity. For example, a restaurant may regard its building as a fixed factor over a period of at least the next year. It would take at least that much time to find a new building or to expand or reduce the size of its present facility. Decisions concerning the operation of the restaurant during the next year must assume the building will remain unchanged. Other factors of production could be changed during the year, but the size of the building must be regarded as a constant.

When the quantity of a factor of production cannot be changed during a particular period, it is called a fixed factor of production. For the restaurant, its building is a fixed factor of production for at least a year. A factor of production whose quantity can be changed during a particular period is called a variable factor of production; factors such as labor and food are examples.

While the managers of the restaurant are making choices concerning its operation over the next year, they are also planning for longer periods. Over those periods, managers may contemplate alternatives such as modifying the building, building a new facility, or selling the building and leaving the restaurant business. The planning period over which a firm can consider *all* factors of production as variable is called the long run.

At any one time, a firm will be making both short-run and long-run choices. The managers may be planning what to do for the next few weeks and for the next few years. Their decisions over the next few weeks are likely to be short-run choices. Decisions that will affect operations over the next few years may be long-run choices, in which managers can consider changing every aspect of their operations.

160. Reading: Short Run and Long Run Average Total Costs

Short Run and Long Run Average Total Costs

As in the short run, costs in the long run depend on the firm's level of output, the costs of factors, and the quantities of factors needed for each level of output. The chief difference between long- and short-run costs is there are no fixed factors in the long run. There are thus no fixed costs. All costs are variable, so we do not distinguish between total variable cost and total cost in the long run: total cost *is* total variable cost.

The long-run average cost (*LRAC*) curve shows the firm's lowest cost per unit at each level of output, assuming that all factors of production are variable. The *LRAC* curve assumes that the firm has chosen the optimal factor mix, as described in the previous section, for producing any level of output. The costs it shows are therefore the lowest costs possible for each level of output. It is important to note, however, that this does not mean that the minimum points of each short-run ATC curves lie on the *LRAC* curve. This critical point is explained in the next paragraph and expanded upon even further in the next section.

Figure 8.9 "Relationship Between Short-Run and Long-Run Average Total Costs" shows how a firm's *LRAC* curve is derived. Suppose Lifetime Disc Co. produces compact discs (CDs) using capital and labor. We have already seen how a firm's average total cost curve can be drawn in the short run for a given quantity of a particular factor of production, such as capital. In the short run, Lifetime Disc might be limited to operating with a given amount of capital; it would face one of the short-run average total cost curves shown in Figure 8.9 “Relationship Between Short-Run and Long-Run Average Total Costs.” If it has 30 units of capital, for example, its average total cost curve is ATC_{30}. In the long run the firm can examine the average total cost curves associated with varying levels of capital.

Four possible short-run average total cost curves for Lifetime Disc are shown in Figure 8.9 “Relationship Between Short-Run and Long-Run Average Total Costs” for quantities of capital of 20, 30, 40, and 50 units. The relevant curves are labeled ATC_{20}, ATC_{30}, ATC_{40}, and ATC_{50} respectively. The *LRAC* curve is derived from this set of short-run curves by finding the lowest average total cost associated with each level of output. Again, notice that the U-shaped *LRAC* curve is an envelope curve that surrounds the various short-run ATC curves. With the exception of ATC_{40}, in this example, the lowest cost per unit for a particular level of output in the long run is not the minimum point of the relevant short-run curve.

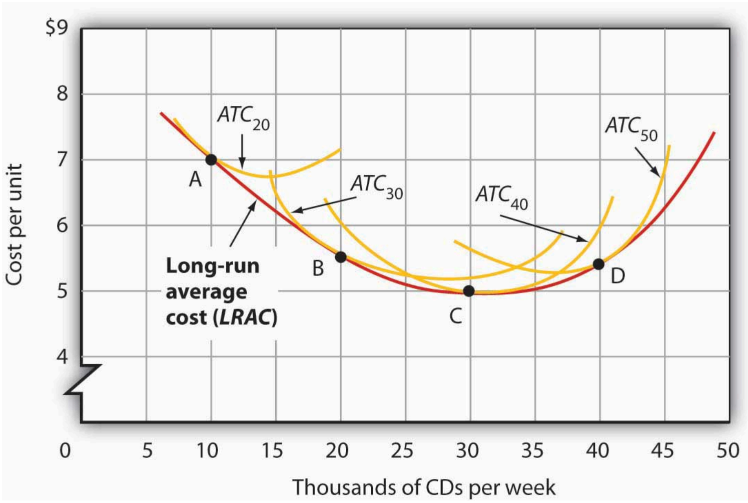


Figure 8.9 Relationship Between Short-Run and Long-Run Average Total Costs. *The LRAC curve is found by taking the lowest average total cost curve at each level of output. Here, average total cost curves for quantities of capital of 20, 30, 40, and 50 units are shown for the Lifetime Disc Co. At a production level of 10,000 CDs per week, Lifetime minimizes its cost per CD by producing with 20 units of capital (point A). At 20,000 CDs per week, an expansion to a plant size associated with 30 units of capital minimizes cost per unit (point B). The lowest cost per unit is achieved with production of 30,000 CDs per week using 40 units of capital (point C). If Lifetime chooses to produce 40,000 CDs per week, it will do so most cheaply with 50 units of capital (point D).*

Self Check: The Short Run vs. The Long Run

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/764

161. Outcome: Long Run Costs and Economies of Scale

What you'll learn to do: define and explain long-run costs, economies of scale, diseconomies of scale, and constant returns to scale

Have you ever wondered why toilet paper at Costco seems so much cheaper there than at your local grocery story? In this outcome, you'll how Costco and other warehouse stores manage to keep costs so low. You'll learn about economies of scale and why it's helpful to think of costs in the long-run.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Long Run Costs
- Reading: Economies of Scale
- Self Check: Long Run Costs and Economies of Scale

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

162. Reading: Long Run Costs

Long Run Costs

The long run is the period of time when all costs are variable. The long run depends on the specifics of the firm in question—it is not a precise period of time. If you have a one-year lease on your factory, then the long run is any period longer than a year, since after a year you are no longer bound by the lease. No costs are fixed in the long run. A firm can build new factories and purchase new machinery, or it can close existing facilities. In planning for the long run, the firm will compare alternative *production technologies* (or processes).

In this context, technology refers to all alternative methods of combining inputs to produce outputs. It does not refer to a specific new invention like the tablet computer. The firm will search for the production technology that allows it to produce the desired level of output at the lowest cost. After all, lower costs lead to higher profits—at least if total revenues remain unchanged. Moreover, each firm must fear that if it does not seek out the lowest-cost methods of production, then it may lose sales to competitor firms that find a way to produce and sell for less.

Choice of Production Technology

Many tasks can be performed with a range of combinations of labor and physical capital. For example, a firm can have human beings answering phones and taking messages, or it can invest in an automated voicemail system. A firm can hire file clerks and secretaries to manage a system of paper folders and file cabinets, or it can invest in a computerized record keeping system that will require fewer employees. A firm can hire workers to push supplies around a factory on rolling carts, it can invest in motorized vehicles, or it can invest in robots that carry materials without a driver. Firms often face a choice between buying many small machines, which need a worker to run each one, or buying one larger and more expensive machine, which requires only one or two workers to operate it. In short, physical capital and labor can often substitute for each other.

Consider the example of a private firm that is hired by local governments to clean up public parks. Three different combinations of labor and physical capital for cleaning up a single average-sized park appear in Table 7.4. The first production technology is heavy on workers and light on machines, while the next two technologies substitute machines for workers. Since all three of these production methods produce the same thing—one cleaned-up park—a profit-seeking firm will choose the production technology that is least expensive, given the prices of labor and machines.

Table 7.4 Three Ways to Clean a Park

Production technology 1	10 workers	2 machines
Production technology 2	7 workers	4 machines
Production technology 3	3 workers	7 machines

Production technology 1 uses the most labor and least machinery, while production technology 3 uses the least labor and the most machinery. Table 7.5 outlines three examples of how the total cost will change with each production technology as the cost of labor changes. As the cost of labor rises from example A to B to C, the firm will choose to substitute away from labor and use more machinery.

Table 7.5 Total Cost with Rising Labor Costs

Example A: Workers cost $40, machines cost $80			
	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	10 × $40 = $400	2 × $80 = $160	$560
Cost of technology 2	7 × $40 = $280	4 × $80 = $320	$600
Cost of technology 3	3 × $40 = $120	7 × $80 = $560	$680
Example B: Workers cost $55, machines cost $80			
	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	10 × $55 = $550	2 × $80 = $160	$710
Cost of technology 2	7 × $55 = $385	4 × $80 = $320	$705
Cost of technology 3	3 × $55 = $165	7 × $80 = $560	$725
Example C: Workers cost $90, machines cost $80			
	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	10 × $90 = $900	2 × $80 = $160	$1,060
Cost of technology 2	7 × $90 = $630	4 × $80 = $320	$950
Cost of technology 3	3 × $90 = $270	7 × $80 = $560	$830

Example A shows the firm's cost calculation when wages are $40 and machines costs are $80. In this case, technology 1 is the low-cost production technology. In example B, wages rise to $55, while the cost of machines does not change, in which case technology 2 is the low-cost production technology. If wages keep rising up to $90, while the cost of machines remains unchanged, then technology 3 clearly becomes the low-cost form of production, as shown in example C.

This example shows that as an input becomes more expensive (in this case, the labor input), firms will attempt to conserve on using that input and will instead shift to other inputs that are relatively less expensive. This pattern helps to explain why the demand curve for labor (or any input) slopes down; that is, as labor becomes relatively more expensive, profit-seeking firms will seek to substitute the use of other inputs. When a multinational employer like Coca-Cola or McDonald's sets up a bottling plant or a restaurant in a high-wage economy like the United States, Canada, Japan, or

Western Europe, it is likely to use production technologies that conserve on the number of workers and focuses more on machines. However, that same employer is likely to use production technologies with more workers and less machinery when producing in a lower-wage country like Mexico, China, or South Africa.

163. Reading: Economies of Scale

Economies of Scale

Once a firm has determined the least costly production technology, it can consider the optimal scale of production, or quantity of output to produce. Many industries experience economies of scale. ***Economies of scale*** refers to the situation where, as the quantity of output goes up, the cost per unit goes down. This is the idea behind "warehouse stores" like Costco or Walmart. In everyday language: a larger factory can produce at a lower average cost than a smaller factory. Figure 7.5 illustrates the idea of economies of scale, showing the average cost of producing an alarm clock falling as the quantity of output rises. For a small-sized factory like S, with an output level of 1,000, the average cost of production is $12 per alarm clock. For a medium-sized factory like M, with an output level of 2,000, the average cost of production falls to $8 per alarm clock. For a large factory like L, with an output of 5,000, the average cost of production declines still further to $4 per alarm clock.

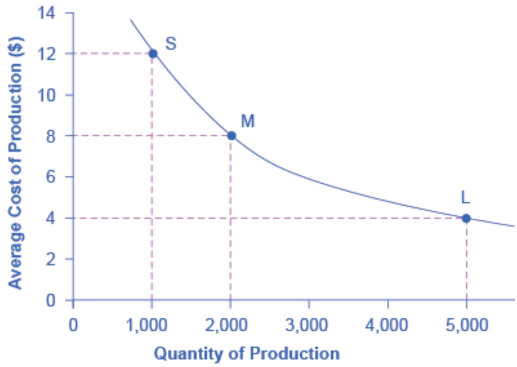


Figure 7.5. Economies of Scale *A small factory like S produces 1,000 alarm clocks at an average cost of $12 per clock. A medium factory like M produces 2,000 alarm clocks at a cost of $8 per clock. A large factory like L produces 5,000 alarm clocks at a cost of $4 per clock. Economies of scale exist because the larger scale of production leads to lower average costs.*

The average cost curve in Figure 7.5 may appear similar to the average cost curves presented earlier in this module, although it is downward-sloping rather than U-shaped. But there is one major difference. The economies of scale curve is a long-run average cost curve, because it allows all factors of production to change. The short-run average cost curves presented earlier in this module assumed the existence of fixed costs, and only variable costs were allowed to change. One prominent example of economies of scale occurs in the chemical industry. Chemical plants have a lot of pipes. The cost of the materials for producing a pipe is related to the circumference of the pipe and its length. However, the volume of chemicals that can flow through a pipe is determined by the cross-section area of the pipe. The calculations in Table 7.6 show that a pipe which uses twice as much material to make (as shown by the circumference of the pipe doubling) can actually carry four times the volume of chemicals because the cross-section area of the pipe rises by a factor of four (as shown in the Area column).

Table 7.6 Comparing Pipes: Economies of Scale in the Chemical Industry

	Circumference (2πr)	Area (πr2)
4-inch pipe	12.5 inches	12.5 square inches
8-inch pipe	25.1 inches	50.2 square inches
16-inch pipe	50.2 inches	201.1 square inches

A doubling of the cost of producing the pipe allows the chemical firm to process four times as much material. This pattern is a major reason for economies of scale in chemical production, which uses a large quantity of pipes. Of course, economies of scale in a chemical plant are more complex than this simple calculation suggests. But the chemical engineers who design these plants have long used what they call the "six-tenths rule," a rule of thumb which holds that increasing the quantity produced in a chemical plant by a certain percentage will increase total cost by only six-tenths as much.

Shapes of Long-Run Average Cost Curves

While in the short run firms are limited to operating on a single average cost curve (corresponding to the level of fixed costs they have chosen), in the long run when all costs are variable, they can choose to operate on any average cost curve. Thus, the ***long-run average cost (LRAC) curve*** is actually based on a group of ***short-run average cost (SRAC) curves***, each of which represents one specific level of fixed costs. More precisely, the long-run average cost curve will be the least expensive average cost curve for any level of output. Figure 7.6 shows how the long-run average cost curve is built from a group of short-run average cost curves. Five short-run-average cost curves appear on the diagram. Each SRAC curve represents a different level of fixed costs. For example, you can imagine SRAC1 as a small factory, SRAC2 as a medium factory, SRAC3 as a large factory, and SRAC4 and SRAC5 as very large and ultra-large. Although this diagram shows only five SRAC curves, presumably there are an infinite number of other SRAC curves between the ones that are shown. This family of short-run average cost curves can be thought of as representing different choices for a firm that is planning its level of investment in fixed cost physical capital—knowing that different choices about capital investment in the present will cause it to end up with different short-run average cost curves in the future.

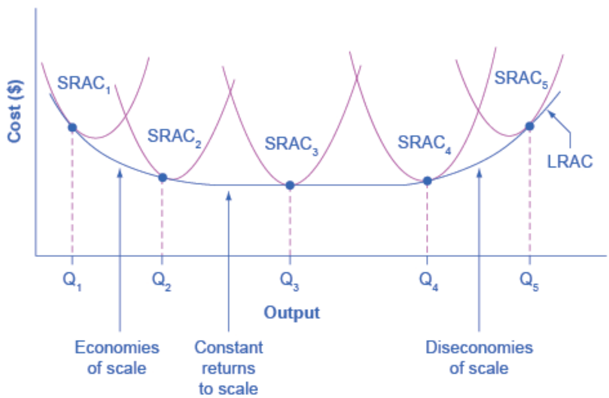


Figure 7.6. From Short-Run Average Cost Curves to Long-Run Average Cost Curves *The five different short-run average cost (SRAC) curves each represents a different level of fixed costs, from the low level of fixed costs at SRAC1 to the high level of fixed costs at SRAC5. Other SRAC curves, not shown in the diagram, lie between the ones that are shown here. The long-run average cost (LRAC) curve shows the lowest cost for producing each quantity of output when fixed costs can vary, and so it is formed by the bottom edge of the family of SRAC curves. If a firm wished to produce quantity Q3, it would choose the fixed costs associated with SRAC3.*

The long-run average cost curve shows the cost of producing each quantity in the long run, when the firm can choose its level of fixed costs and thus choose which short-run average costs it desires. If the firm plans to produce in the long run at an output of

Q3, it should make the set of investments that will lead it to locate on SRAC3, which allows producing q3 at the lowest cost. A firm that intends to produce Q3 would be foolish to choose the level of fixed costs at SRAC2 or SRAC4. At SRAC2 the level of fixed costs is too low for producing Q3 at lowest possible cost, and producing q3 would require adding a very high level of variable costs and make the average cost very high. At SRAC4, the level of fixed costs is too high for producing q3 at lowest possible cost, and again average costs would be very high as a result. The shape of the long-run cost curve, as drawn in Figure 7.6, is fairly common for many industries. The left-hand portion of the long-run average cost curve, where it is downward-sloping from output levels Q1 to Q2 to Q3, illustrates the case of economies of scale. In this portion of the long-run average cost curve, larger scale leads to lower average costs. This pattern was illustrated earlier in Figure 7.5. In the middle portion of the long-run average cost curve, the flat portion of the curve around Q3, economies of scale have been exhausted. In this situation, allowing all inputs to expand does not much change the average cost of production, and it is called *constant returns to scale*. In this range of the LRAC curve, the average cost of production does not change much as scale rises or falls. The following feature explains where diminishing marginal returns fit into this analysis.

How do Economies of Scale Compare to Diminishing Marginal Returns?

The concept of economies of scale, where average costs decline as production expands, might seem to conflict with the idea of diminishing marginal returns, where marginal costs rise as production expands. But diminishing marginal returns refers only to the short-run average cost curve, where one variable input (like labor) is increasing, but other inputs (like capital) are fixed. Economies of scale refers to the long-run average cost curve where all inputs are being allowed to increase together. Thus, it is quite possible and common to have an industry that has both diminishing marginal returns when only one input is allowed to change, and at the same time has increasing or constant economies of scale when all inputs change together to produce a larger-scale operation.

Finally, the right-hand portion of the long-run average cost curve, running from output level Q4 to Q5, shows a situation where, as the level of output and the scale rises, average costs rise as well. This situation is called *diseconomies of scale*. A firm or a factory can grow so large that it becomes very difficult to manage, resulting in unnecessarily high costs as many layers of management try to communicate with workers and with each other, and as failures to communicate lead to disruptions in the flow of work and materials. Not many overly large factories exist in the real world, because with their very high production costs, they are unable to compete for long against plants with lower average costs of production. However, in some planned economies, like the economy of the old Soviet Union, plants that were so large as to be grossly inefficient were able to continue operating for a long time because government economic planners protected them from competition and ensured that they would not make losses. Diseconomies of scale can also be present across an entire firm, not just a large factory. The *leviathan effect* can hit firms that become too large to run efficiently, across the entirety of the enterprise. Firms that shrink their operations are often responding to finding itself in the diseconomies region, thus moving back to a lower average cost at a lower output level.

LINK IT UP

Visit this website to read an article about the complexity of the belief that banks can be "too-big-to-fail."

The Size and Number of Firms in an Industry

The shape of the long-run average cost curve has implications for how many firms will compete in an industry, and whether the firms in an industry have many different sizes, or tend to be the same size. For example, say that one million dishwashers are sold every year at a price of $500 each and the long-run average cost curve for dishwashers is shown in Figure 7.7 (a). In Figure 7.7 (a), the lowest point of the LRAC curve occurs at a quantity of 10,000 produced. Thus, the market for dishwashers will consist of 100 different manufacturing plants of this same size. If some firms built a plant that produced 5,000 dishwashers per year or 25,000 dishwashers per year, the average costs of production at such plants would be well above $500, and the firms would not be able to compete.

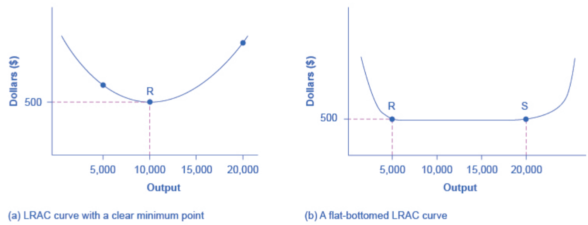


Figure 7.7. The LRAC Curve and the Size and Number of Firms *(a) Low-cost firms will produce at output level R. When the LRAC curve has a clear minimum point, then any firm producing a different quantity will have higher costs. In this case, a firm producing at a quantity of 10,000 will produce at a lower average cost than a firm producing, say, 5,000 or 20,000 units. (b) Low-cost firms will produce between output levels R and S. When the LRAC curve has a flat bottom, then firms producing at any quantity along this flat bottom can compete. In this case, any firm producing a*

quantity between 5,000 *and* 20,000 *can compete effectively, although firms producing less than* 5,000 *or more than* 20,000 *would face higher average costs and be unable to compete.*

How Can Cities be Viewed As Examples of Economies of Scale?

Why are people and economic activity concentrated in cities, rather than distributed evenly across a country? The fundamental reason must be related to the idea of economies of scale—that grouping economic activity is more productive in many cases than spreading it out. For example, cities provide a large group of nearby customers, so that businesses can produce at an efficient economy of scale. They also provide a large group of workers and suppliers, so that business can hire easily and purchase whatever specialized inputs they need. Many of the attractions of cities, like sports stadiums and museums, can operate only if they can draw on a large nearby population base. Cities are big enough to offer a wide variety of products, which is what many shoppers are looking for.

These factors are not exactly economies of scale in the narrow sense of the production function of a single firm, but they are related to growth in the overall size of population and market in an area. Cities are sometimes called "agglomeration economies."

These agglomeration factors help to explain why every economy, as it develops, has an increasing proportion of its population living in urban areas. In the United States, about 80% of the population now lives in metropolitan areas (which include the suburbs around cities), compared to just 40% in 1900. However, in poorer nations of the world, including much of Africa, the proportion of the population in urban areas is only about 30%. One of the great challenges for these countries as their economies grow will be to manage the growth of the great cities that will arise.

If cities offer economic advantages that are a form of economies of scale, then why don't all or most people live in one giant city? At some point, agglomeration economies must turn into diseconomies. For example, traffic congestion may reach a point where the gains from being geographically nearby are counterbalanced by how long it takes to travel. High densities of people, cars, and factories can mean more garbage and air and water pollution. Facilities like parks or museums may become overcrowded. There may be economies of scale for negative activities like crime, because high densities of people and businesses, combined with the greater impersonality of cities, make it easier for illegal activities as well as legal ones. The future of cities, both in the United States and in other countries around the world, will be determined by their ability to benefit from the economies of agglomeration and to minimize or counterbalance the corresponding diseconomies.

A more common case is illustrated in Figure 7.7 (b), where the LRAC curve has a flat-bottomed area of constant returns to scale. In this situation, any firm with a level of output between 5,000 and 20,000 will be able to produce at about the same level of average cost. Given that the market will demand one million dishwashers per year at a price of $500, this market might have as many as 200 producers (that is, one million dishwashers divided by firms making 5,000 each) or as few as 50 producers (one million dishwashers divided by firms making 20,000 each). The producers in this market will range in size from firms that make 5,000 units to firms that make 20,000 units. But firms that produce below 5,000 units or more than 20,000 will be unable to compete, because their average costs will be too high. Thus, if we see an industry where almost all plants are the same size, it is likely that the long-run average cost curve has a unique bottom point as in Figure 7.7 (a). However, if the long-run average cost curve has a wide flat bottom like Figure 7.7 (b), then firms of a variety of different sizes will be able to compete with each other.

The flat section of the long-run average cost curve in Figure 7.7

(b) can be interpreted in two different ways. One interpretation is that a single manufacturing plant producing a quantity of 5,000 has the same average costs as a single manufacturing plant with four times as much capacity that produces a quantity of 20,000. The other interpretation is that one firm owns a single manufacturing plant that produces a quantity of 5,000, while another firm owns four separate manufacturing plants, which each produce a quantity of 5,000. This second explanation, based on the insight that a single firm may own a number of different manufacturing plants, is especially useful in explaining why the long-run average cost curve often has a large flat segment—and thus why a seemingly smaller firm may be able to compete quite well with a larger firm. At some point, however, the task of coordinating and managing many different plants raises the cost of production sharply, and the long-run average cost curve slopes up as a result.

In the examples to this point, the quantity demanded in the market is quite large (one million) compared with the quantity produced at the bottom of the long-run average cost curve (5,000, 10,000 or 20,000). In such a situation, the market is set for competition between many firms. But what if the bottom of the long-run average cost curve is at a quantity of 10,000 and the total market demand at that price is only slightly higher than that quantity—or even somewhat lower?

Return to Figure 7.7 (a), where the bottom of the long-run average cost curve is at 10,000, but now imagine that the total quantity of dishwashers demanded in the market at that price of $500 is only 30,000. In this situation, the total number of firms in the market would be three. A handful of firms in a market is called an "oligopoly," and the module on Monopolistic Competition and Oligopoly will discuss the range of competitive strategies that can occur when oligopolies compete.

Alternatively, consider a situation, again in the setting of Figure 7.7 (a), where the bottom of the long-run average cost curve is 10,000, but total demand for the product is only 5,000. (For simplicity, imagine that this demand is highly inelastic, so that it does not vary according to price.) In this situation, the market may well end up with a single firm—a monopoly—producing all 5,000 units. If any firm tried to challenge this monopoly while producing a quantity lower than 5,000 units, the prospective competitor firm would have a higher average cost, and so it would not be able to compete in the longer term without losing money. The module on Monopoly discusses the situation of a monopoly firm.

Thus, the shape of the long-run average cost curve reveals whether competitors in the market will be different sizes. If the LRAC curve has a single point at the bottom, then the firms in the market will be about the same size, but if the LRAC curve has a flat-bottomed segment of constant returns to scale, then firms in the market may be a variety of different sizes. The relationship between the quantity at the minimum of the long-run average cost curve and the quantity demanded in the market at that price will predict how much competition is likely to exist in the market. If the quantity demanded in the market far exceeds the quantity at the minimum of the LRAC, then many firms will compete. If the quantity demanded in the market is only slightly higher than the quantity at the minimum of the LRAC, a few firms will compete. If the quantity demanded in the market is less than the quantity at the minimum of the LRAC, a single-producer monopoly is a likely outcome.

Shifting Patterns of Long-Run Average Cost

New developments in production technology can shift the long-run average cost curve in ways that can alter the size distribution of firms in an industry.

For much of the twentieth century, the most common change has been to see alterations in technology, like the assembly line or the large department store, where large-scale producers seemed to gain an advantage over smaller ones. In the long-run average cost curve, the downward-sloping economies of scale portion of the curve stretched over a larger quantity of output.

However, new production technologies do not inevitably lead to a greater average size for firms. For example, in recent years some new technologies for generating electricity on a smaller scale have appeared. The traditional coal-burning electricity plants needed to produce 300 to 600 megawatts of power to exploit economies of scale fully. However, high-efficiency turbines to produce electricity from burning natural gas can produce electricity at a competitive price while producing a smaller quantity of 100 megawatts or less. These new technologies create the possibility for smaller companies or plants to generate electricity as efficiently as large ones. Another example of a technology-driven shift to smaller plants may be taking place in the tire industry. A traditional mid-size tire plant produces about six million tires per year. However, in 2000, the Italian company Pirelli introduced a new tire factory that uses many robots. The Pirelli tire plant produced only about one million tires per year, but did so at a lower average cost than a traditional mid-sized tire plant.

Controversy has simmered in recent years over whether the new information and communications technologies will lead to a larger or smaller size for firms. On one side, the new technology may make it easier for small firms to reach out beyond their local geographic area and find customers across a state, or the nation, or even across international boundaries. This factor might seem to predict a future with a larger number of small competitors. On the other side, perhaps the new information and communications technology will create "winner-take-all" markets where one large company will tend to command a large share of total sales, as Microsoft has done in the production of software for personal computers or Amazon has done in online bookselling. Moreover, improved information and communication technologies might make it easier to manage many different plants and operations across the country or around the world, and thus encourage larger firms. This ongoing battle between the forces of smallness and largeness will be of great interest to economists, businesspeople, and policymakers.

Amazon

Traditionally, bookstores have operated in retail locations with inventories held either on the shelves or in the back of the store. These retail locations were very pricey in terms of rent. Amazon has no retail locations; it sells online and delivers by mail. Amazon offers almost any book in print, convenient purchasing, and prompt delivery by mail. Amazon holds its inventories in huge warehouses in low-rent locations around the world. The warehouses are highly computerized using robots and relatively low-skilled workers, making for low average costs per sale. Amazon demonstrates the significant advantages economies of scale can offer to a firm that exploits those economies.

Self Check: Long-Run Costs and Economies of Scale

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the three Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/650

164. Putting It Together: Production

Summary

The goal of this module was to explore the relationship between the inputs used in production, the resulting output and the cost of that output. You learned how to:

- Define the term "production" and explain what a production function is
- Define and differentiate between marginal, average, and total product; compute and graph marginal, average, and total product
- Differentiate between explicit and implicit costs, accounting and economic profit
- Define and differentiate between marginal, average, fixed, variable and total costs; compute and graph marginal, average, fixed, variable and total costs
- Differentiate between short-run and long-run costs
- Define and explain long-run costs

Examples

This module was detailed and complicated, but that's because production and cost issues are too. There are a near infinite number of different types of businesses and many have a unique production process. The local bakery down the street and Wonder bread don't operate the same way or at the same scale. Neither does a home baker. So each production process and costs of production are bound to be different.

Bread Collage. [Memphis Wonder Bread Factory] by Mr. Littlehand, [CC-BY]. [Corner Bakery] by Guiseppi Milo, [CC-BY]. [Home Baker] Cheryl, [CC-BY-SA]

What should you remember from this module?

- Any change in the production process will cause a change in production cost.
- Average and marginal costs are per unit; Fixed, variable and total costs are just dollars spent on all units of output produced.
- Long-run costs are usually less than short-run costs because you have more options.
- Scale matters for some industries, but not others. The cost of producing electricity per kilowatt-hour is much less than the cost of using a home generator. But the cost of running a taxi business with 100 cabs is not much different than running the same business with 25 cabs.

Let's return to the question posed at the beginning of the module: How does an improvement in technology affect production and costs? The answer is simple. An improvement in technology usually reduces the cost of producing a given quantity of output. In other words, a firm can produce the same amount of output at lower cost or more output for the same total cost. This can be shown graphically by an upward shift in the production curves (showing more Q for the same L) and by a downward (or rightward) shift in the cost curves (showing less cost for the same Q). Make sense?

165. Glossary: Production

accounting profit
total revenues minus explicit costs, including depreciation

average profit
profit divided by the quantity of output produced; profit margin

average total cost
total cost divided by the quantity of output

average variable cost
variable cost divided by the quantity of output

constant returns to scale
expanding all inputs proportionately does not change the average cost of production

diseconomies of scale
the long-run average cost of producing each individual unit increases as total output increases

economic profit
total revenues minus total costs (explicit plus implicit costs)

explicit costs
out-of-pocket costs for a firm, for example, payments for wages and salaries, rent, or materials

firm
an organization that combines inputs of labor, capital, land, and raw or finished component materials to produce outputs.

fixed cost
expenditure that must be made before production starts and that does not change regardless of the level of production

implicit costs
: opportunity cost of resources already owned by the firm and used in business, for example, expanding a factory onto land already owned

long-run average cost (LRAC) curve
: shows the lowest possible average cost of production, allowing all the inputs to production to vary so that the firm is choosing its production technology

marginal cost
: the additional cost of producing one more unit

private enterprise
: the ownership of businesses by private individuals

production technologies
: alternative methods of combining inputs to produce output

production
: the process of combining inputs to produce outputs, ideally of a value greater than the value of the inputs

revenue
: income from selling a firm's product; defined as price times quantity sold

short-run average cost (SRAC) curve
: the average total cost curve in the short term; shows the total of the average fixed costs and the average variable costs

total cost
: the sum of fixed and variable costs of production

variable cost
: cost of production that increases with the quantity produced

166. Discussion: Diminishing Returns

What do economists mean by “diminishing returns” to an input? What causes diminishing returns? Have you ever observed this principle at work in a job you’ve had? Describe how you’ve experienced this concept in the real world.

PART X

MODULE: PERFECT COMPETITION

167. Why It Matters: Perfect Competition

Why analyze a firm's profit maximizing decisions under conditions of perfect competition?

This module is the second in the theory of the firm and the first of four modules examining models of market structure. Market structure means, in a nutshell, how competitive or monopolistic is a particular industry. It should be clear that market structure influences how firms behave.

We start by looking at the ideal model of perfect competition. This model is a bit of a head scratcher since there are, actually, very few examples of industries like this in the real world. Why then do we study it? Here's a question for you to think about as you move through the module: What's so perfect about perfect competition? Hint: the model has certain ideal features that you will learn.

Tomato Collage. Farmer's Market by Andre Natta, CC-BY. Firey Tomatoes by Swaminathan, CC-BY. NorthCharleston Farmer's Market by NorthCharleston, CC-BY-SA. What's Better than Love CC torbakhopper, CC-BY.

Have you ever noticed that all the tomatoes of the same type in a farmer's market cost about the same price? The same thing is true of roadside vegetable stands in the countryside. If one stall in a locality has tomatoes for $3 per pound, they all do. Now the price may change from week to week, but it's always the same across the different vendors in the market. You will soon learn why this is.

There are more similarities than differences between this and the following three modules. What you learn in this module will carry over and help you understand the next ones, so the more effort you put into learning this one, the easier the next three modules will be.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=194

LEARNING OUTCOMES

- Define the characteristics of Perfect Competition
- Understand the difference between the firm and the industry
- Calculate and graph the firm's fixed, variable, average, marginal and total costs
- Measure variable and total costs as the area under the average variable and average total cost curves
- Determine the break-even, and the shutdown points of production for a perfectly competitive firm
- Explain the difference between short run and long run equilibrium
- Understand why perfectly competitive markets are efficient

168. Outcome: Defining Perfect Competition

What you'll learn to do: define the characteristics of perfect competition

Imagine the 7-year old you had a lemonade stand. It was one of several on the street. Your neighbor, Julie, also had a lemonade stand and she typically sold her lemonade for 25 cents. You figured that in order to make more money, you would charge 50 cents and steal all her customers. Sadly, everyone bought from Julie and you had no customers at all.

Welcome to the world of perfect competition. You will see in this outcome that because your lemonade stands were essentially identical, in order to remain in business and make any profit, you needed to be a price-taker instead of a price-maker.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Perfect Competition
- Self Check: Defining Perfect Competition

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

169. Reading: Perfect Competition

A Dime a Dozen

When you were younger did you babysit, deliver papers, or mow lawn(s) for money? If so, you faced stiff competition from other competitors who offered identical services. There was nothing to stop others from offering their services too. All of you charged the "going rate." If you tried to charge more, your customers would simply buy from someone else. These conditions are very similar to the conditions agricultural growers face.

Figure 8.1. *Depending upon the competition and prices offered, a wheat farmer may choose to grow a different crop. (Credit: modification of work by Daniel X. O'Neil/Flickr Creative Commons)*

Growing a crop may be more difficult to start than a babysitting or lawn mowing service, but growers face the same fierce competition. In the grand scale of world agriculture, farmers face competition from thousands of others because they sell an identical product. After all, winter wheat is winter wheat. But it is relatively easy for farmers to leave the marketplace for another crop. In this case, they do not sell the family farm, they switch crops.

Take the case of the upper Midwest region of the United States—for many generations the area was called "King Wheat." According to the United States Department of Agriculture National Agricultural Statistics Service, statistics by state, in 1997, 11.6 million acres of wheat and 780,000 acres of corn were planted in North Dakota. In the intervening 15 or so years has the mix of crops changed? Since it is relatively easy to switch crops, did farmers change what was planted as the relative crop prices changed? We will find out at module's end.

In the meantime, let's consider the topic of this module—the perfectly competitive market. This is a market in which entry and exit are relatively easy and competitors are "a dime a dozen."

Introduction to Perfect Competition

All businesses face two realities: no one is required to buy their products, and even customers who might want those products may buy from other businesses instead. Firms that operate in perfectly competitive markets face this reality. In this module you will learn how such firms make decisions about how much to produce, how much profit they make, whether to stay in business or not, and many others. Industries differ from one another in terms of how many sellers there are in a specific market, how easy or difficult it is for a new firm to enter, and the type of products that are sold. This is referred to as the *market structure* of the industry. In this module we focus on perfect competition. However, in other modules we will examine other industry types: Monopoly and Monopolistic Competition and Oligopoly.

Perfect Competition and Why It Matters

Firms are said to be in ***perfect competition*** when the following conditions occur: (1) many firms produce identical products; (2) many buyers are available to buy the product, and many sellers are available to sell the product; (3) sellers and buyers have all relevant information to make rational decisions about the product being bought and sold; and (4) firms can enter and leave the market without any restrictions—in other words, there is free entry and exit into and out of the market.

A ***perfectly competitive firm*** is known as a ***price taker***, because the pressure of competing firms forces them to accept the prevailing equilibrium price in the market. If a firm in a perfectly competitive market raises the price of its product by so much as a penny, it will lose all of its sales to competitors. When a wheat grower wants to know what the going price of wheat is, he or she has to go to the computer or listen to the radio to check. The market price is determined solely by supply and demand in the entire market and not the individual farmer. Also, a perfectly competitive firm must be a very small player in the overall market, so that it can increase or decrease output without noticeably affecting the overall quantity supplied and price in the market.

A perfectly competitive market is a hypothetical extreme; however, producers in a number of industries do face many competitor firms selling highly similar goods, in which case they must often act as price takers. Agricultural markets are often used as an example. The same crops grown by different farmers are largely interchangeable. According to the United States Department of Agriculture monthly reports, in 2012, U.S. corn farmers received an average price of $6.07 per bushel and wheat farmers received an average price of $7.60 per bushel. A corn farmer who attempted to sell at $7.00 per bushel, or a wheat grower who attempted to sell for $8.00 per bushel, would not have found any buyers. A perfectly competitive firm will not sell below the equilibrium price either.

Why should they when they can sell all they want at the higher price? Other examples of agricultural markets that operate in close to perfectly competitive markets are small roadside produce markets and small organic farmers.

LINK IT UP

Visit this website that reveals the current value of various commodities.

This module examines how profit-seeking firms decide how much to produce in perfectly competitive markets. Such firms will analyze their costs as discussed in the module on Cost and Industry Structure. In the short run, the perfectly competitive firm will seek the quantity of output where profits are highest or, if profits are not possible, where losses are lowest. In this example, the "short run" refers to a situation in which firms are producing with one fixed input and incur fixed costs of production. (In the real world, firms can have many fixed inputs.)

In the long run, perfectly competitive firms will react to profits by increasing production. They will respond to losses by reducing production or exiting the market. Ultimately, a long-run *equilibrium* will be attained when no new firms want to enter the market and existing firms do not want to leave the market, as economic profits have been driven down to zero.

Key Concepts and Summary

A perfectly competitive firm is a price taker, which means that it must accept the equilibrium price at which it sells goods. If a perfectly competitive firm attempts to charge even a tiny amount more than the market price, it will be unable to make any sales.

In a perfectly competitive market there are thousands of sellers, easy entry, and identical products. A short-run production period is when firms are producing with some fixed inputs. Long-run equilibrium in a perfectly competitive industry occurs after all firms have entered and exited the industry and seller profits are driven to zero.

Perfect competition means that there are many sellers, there is easy entry and exiting of firms, products are identical from one seller to another, and sellers are price takers.

Self Check: Defining Perfect Competition

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/651

170. Outcome: Perfectly Competitive Firms and Industries

What you'll learn to do: understand the difference between the firm and the industry

The basic idea behind perfect competition is that there are lots of sellers selling almost the same thing, and there are also lots of buyers wanting those products. In this section, you'll understand more about the differences between a perfectly competitive firm and a perfectly competitive industry. While a competitive market determines the equilibrium point by staying in tune with the supply and demand curves, a perfectly competitive market does not have that luxury. A perfectly competitive market must accept the price point and must only decide how much to sell.

The specific things you'll learn to do in this section include:

- Explain and illustrate the differences between the demand curve for a perfectly competitive firm and that for a perfectly competitive industry

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Perfect Competition: A Model
- Reading: Price and Revenue in a Perfectly Competitive Industry

and Firm
- Self Check: Perfectly Competitive Firms and Industries

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

171. Reading: Perfect Competition: A Model

Perfect Competition: A Model

Virtually all firms in a market economy face competition from other firms. In this module, we will be working with a model of a highly idealized form of competition called "perfect" by economists.

Perfect competition is a model of the market based on the assumption that a large number of firms produce identical goods consumed by a large number of buyers. The model of perfect competition also assumes that it is easy for new firms to enter the market and for existing ones to leave. And finally, it assumes that buyers and sellers have complete information about market conditions.

As we examine these assumptions in greater detail, we will see that they allow us to work with the model more easily. No market fully meets the conditions set out in these assumptions. As is always the case with models, our purpose is to understand the way things work, not to describe them. And the model of perfect competition will prove enormously useful in understanding the world of markets.

Assumptions of the Model

The assumptions of the model of perfect competition, taken together, imply that individual buyers and sellers in a perfectly competitive market accept the market price as given. No one buyer or seller has any influence over that price. Individuals or firms who must take the market price as given are called price takers.

A consumer or firm that takes the market price as given has no ability to influence that price. A price-taking firm or consumer is like an individual who is buying or selling stocks. He or she looks up the market price and buys or sells at that price. The price is determined by demand and supply in the market—not by individual buyers or sellers. In a perfectly competitive market, each firm and each consumer is a price taker. A price-taking consumer assumes that he or she can purchase any quantity at the market price—without affecting that price. Similarly, a price-taking firm assumes it can sell whatever quantity it wishes at the market price without affecting the price.

You are a price taker when you go into a store. You observe the prices listed and make a choice to buy or not. Your choice will not affect that price. Should you sell a textbook back to your campus bookstore at the end of a course, you are a price-taking seller. You are confronted by a market price and you decide whether to sell or not. Your decision will not affect that price.

To see how the assumptions of the model of perfect competition imply price-taking behavior, let us examine each of them in turn.

Identical Goods

In a perfectly competitive market for a good or service, one unit of the good or service cannot be differentiated from any other on any basis. A bushel of, say, hard winter wheat is an example. A bushel produced by one farmer is identical to that produced by another. There are no brand preferences or consumer loyalties.

The assumption that goods are identical is necessary if firms are to be price takers. If one farmer's wheat were perceived as having special properties that distinguished it from other wheat, then that farmer would have some power over its price. By assuming that all goods and services produced by firms in a perfectly competitive market are identical, we establish a necessary condition for price-taking behavior. Economists sometimes say that the goods or services in a perfectly competitive market are *homogeneous*, meaning that they are all alike. There are no brand differences in a perfectly competitive market.

A Large Number of Buyers and Sellers

How many buyers and sellers are in our market? The answer rests on our presumption of price-taking behavior. There are so many buyers and sellers that none of them has any influence on the market price regardless of how much any of them purchases or sells. A firm in a perfectly competitive market can react to prices, but cannot affect the prices it pays for the factors of production or the prices it receives for its output.

Ease of Entry and Exit

The assumption that it is easy for other firms to enter a perfectly competitive market implies an even greater degree of competition. Firms in a market must deal not only with the large number of competing firms but also with the possibility that still more firms might enter the market.

Later in this module, we will see how ease of entry is related to the sustainability of economic profits. If entry is easy, then the promise of high economic profits will quickly attract new firms. If entry is difficult, it won't.

The model of perfect competition assumes easy exit as well as easy entry. The assumption of easy exit strengthens the assumption of easy entry. Suppose a firm is considering entering a particular market. Entry may be easy, but suppose that getting out is difficult. For example, suppliers of factors of production to firms in the industry might be happy to accommodate new firms but might require that they sign long-term contracts. Such contracts could make leaving the market difficult and costly. If that were the case, a firm might be hesitant to enter in the first place. Easy exit helps make entry easier.

Complete Information

We assume that all sellers have complete information about prices, technology, and all other knowledge relevant to the operation of the market. No one seller has any information about production methods that is not available to all other sellers. If one seller had an advantage over other sellers, perhaps special information about a lower-cost production method, then that seller could exert some control over market price—the seller would no longer be a price taker.

We assume also that buyers know the prices offered by every seller. If buyers did not know about prices offered by different firms in the market, then a firm might be able to sell a good or service for a price other than the market price and thus could avoid being a price taker.

The availability of information that is assumed in the model of perfect competition implies that information can be obtained at low cost. If consumers and firms can obtain information at low cost, they are likely to do so. Information about the marketplace may come over the internet, over the airways in a television commercial, or over a cup of coffee with a friend. Whatever its source, we assume that its low cost ensures that consumers and firms have enough of it so that everyone buys or sells goods and services at market prices determined by the intersection of demand and supply curves.

The assumptions of the perfectly competitive model ensure that each buyer or seller is a price taker. The market, not individual consumers or firms, determines price in the model of perfect competition. No individual has enough power in a perfectly competitive market to have any impact on that price.

Perfect Competition and the Real World

The assumptions of identical products, a large number of buyers, easy entry and exit, and perfect information are strong assumptions. The notion that firms must sit back and let the market determine price seems to fly in the face of what we know about most real firms, which is that firms customarily do set prices. Yet this is the basis for the model of demand and supply, the power of which you have already seen.

When we use the model of demand and supply, we assume that market forces determine prices. In this model, buyers and sellers respond to the market price. They are price takers. The assumptions of the model of perfect competition underlie the assumption of price-taking behavior. Thus we are using the model of perfect competition whenever we apply the model of demand and supply.

We can understand most markets by applying the model of demand and supply. Even though those markets do not fulfill all the assumptions of the model of perfect competition, the model allows us to understand some key features of these markets.

Changes within your lifetime have made many markets more competitive. Falling costs of transportation, together with dramatic advances in telecommunications, have opened the possibility of entering markets to firms all over the world. A company in South Korea can compete in the market for steel in the United States. A furniture maker in New Mexico can compete in the market for furniture in Japan. A firm can enter the world market simply by creating a web page to advertise its products and to take orders.

In the remaining sections of this module, we will learn more about the response of firms to market prices. We will see how firms respond, in the short run and in the long run, to changes in demand and to changes in production costs. In short, we will be examining the forces that constitute the supply side of the model of demand and supply.

We will also see how competitive markets work to serve consumer interests and how competition acts to push economic profits down, sometimes eliminating them entirely. When we have finished we will have a better understanding of the market conditions facing farmers and of the conditions that prevail in any competitive industry.

Watch this video to learn more about the characteristics of perfect competition:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=198#oembed-1

KEY TAKEAWAYS

- The central characteristic of the model of perfect competition is the fact that price is determined by the interaction of demand and supply; buyers and sellers are price takers.
- The model assumes: a large number of firms producing identical (homogeneous) goods or services, a large number of buyers and sellers, easy entry and exit in the industry, and complete information about prices in the market.
- The model of perfect competition underlies the model of demand and supply.

172. Reading: Price and Revenue in a Perfectly Competitive Industry and Firm

Price and Revenue

Each firm in a perfectly competitive market is a price taker; the equilibrium price and industry output are determined by demand and supply. Figure 9.1 “The Market for Radishes” shows how demand and supply in the market for radishes, which we shall assume are produced under conditions of perfect competition, determine total output and price. The equilibrium price is $0.40 per pound; the equilibrium quantity is 10 million pounds per month.

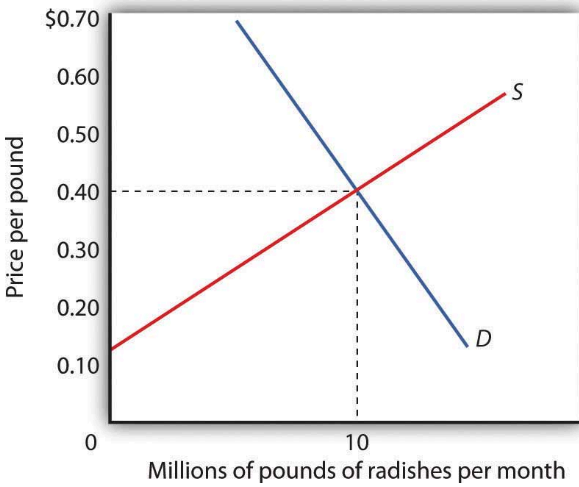


Figure 9.1 *The Market for Radishes. Price and output in a competitive market are determined by demand and supply. In the market for radishes, the equilibrium price is $0.40 per pound; 10 million pounds per month are produced and purchased at this price.*

Because it is a price taker, each firm in the radish industry assumes it can sell all the radishes it wants at a price of $0.40 per pound. No matter how many or how few radishes it produces, the firm expects to sell them all at the market price.

The assumption that the firm expects to sell all the radishes it wants at the market price is crucial. If a firm did not expect to sell all of its radishes at the market price—if it had to lower the price to sell some quantities—the firm would not be a price taker. And price-taking behavior is central to the model of perfect competition.

Radish growers—and perfectly competitive firms in general—have no reason to charge a price lower than the market price. Because buyers have complete information and because we assume each firm's product is identical to that of its rivals, firms are unable to charge a price higher than the market price. For perfectly competitive firms, the price is very much like the weather: they may complain about it, but in perfect competition there is nothing any of them can do about it.

This video explains how the market supply and demand curves determine the price of a good, and why firms in a perfectly competitive market are price takers.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=199#oembed-1

Total Revenue

While a firm in a perfectly competitive market has no influence over its price, it does determine the output it will produce. In selecting the quantity of that output, one important consideration is the revenue the firm will gain by producing it.

A firm's total revenue is found by multiplying its output by the price at which it sells that output. For a perfectly competitive firm, total revenue (*TR*) is the market price (*P*) times the quantity the firm produces (Q), or

$$TR = P \times Q$$

The relationship between market price and the firm's total revenue curve is a crucial one. Panel (a) of Figure 9.2 "Total Revenue, Marginal Revenue, and Average Revenue" shows total revenue curves for a radish grower at three possible market prices: $0.20, $0.40, and $0.60 per pound. Each total revenue curve is a linear, upward-sloping curve. At any price, the greater the quantity a perfectly competitive firm sells, the greater its total revenue. Notice that the greater the price, the steeper the total revenue curve is.

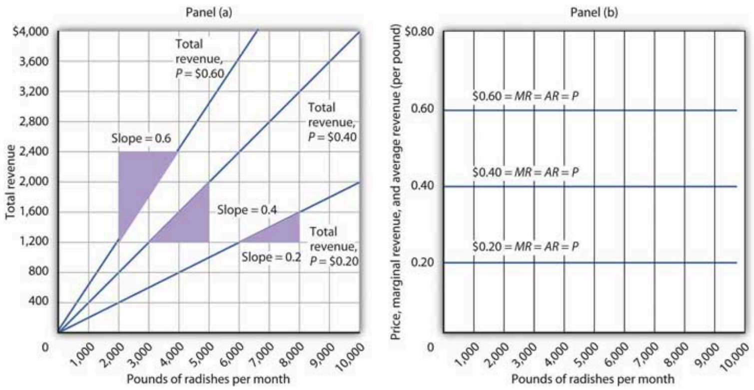


Figure 9.2 *Total Revenue, Marginal Revenue, and Average Revenue. Panel (a) shows different total revenue curves for three possible market prices in perfect competition. A total revenue curve is a straight line coming out of the origin. The slope of a total revenue curve is MR; it equals the market price (P) and AR in perfect competition. Marginal revenue and average revenue are thus a single horizontal line at the market price, as shown in Panel (b). There is a different marginal revenue curve for each price.*

Marginal Revenue, Price, and Demand for the Perfectly Competitive Firm

We have seen that a perfectly competitive firm's marginal revenue curve is simply a horizontal line at the market price and that this same line is also the firm's average revenue curve. For the perfectly competitive firm, MR=P=AR. The marginal revenue curve has another meaning as well. It is the demand curve facing a perfectly competitive firm.

Consider the case of a single radish producer, Tony Gortari. We assume that the radish market is perfectly competitive; Mr. Gortari runs a perfectly competitive firm. Suppose the market price of radishes is $0.40 per pound. How many pounds of radishes can Mr.

Gortari sell at this price? The answer comes from our assumption that he is a price taker: He can sell *any* quantity he wishes at this price. How many pounds of radishes will he sell if he charges a price that exceeds the market price? None. His radishes are identical to those of every other firm in the market, and everyone in the market has complete information. That means the demand curve facing Mr. Gortari is a horizontal line at the market price as illustrated in Figure 9.3 "Price, Marginal Revenue, and Demand". Notice that the curve is labeled *d* to distinguish it from the market demand curve, *D*, in Figure 9.1 "The Market for Radishes". The horizontal line in Figure 9.3 "Price, Marginal Revenue, and Demand" is also Mr. Gortari's marginal revenue curve, *MR*, and his average revenue curve, *AR*. It is also the market price, *P*.

Of course, Mr. Gortari could charge a price below the market price, but why would he? We assume he can sell all the radishes he wants at the market price; there would be no reason to charge a lower price. Mr. Gortari faces a demand curve that is a horizontal line at the market price. In our subsequent analysis, we shall refer to the horizontal line at the market price simply as marginal revenue. We should remember, however, that this same line gives us the market price, average revenue, and the demand curve facing the firm.

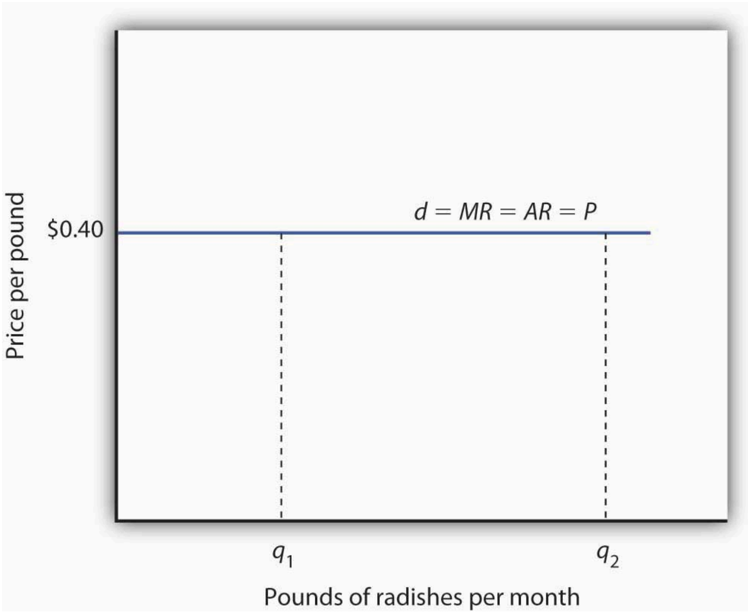


***Figure 9.3** Price, Marginal Revenue, and Demand. A perfectly competitive firm faces a horizontal demand curve at the market price. Here, radish grower Tony Gortari faces demand curve d at the market price of $0.40 per pound. He could sell q1 or q2—or any other quantity—at a price of $0.40 per pound.*

More generally, we can say that *any* perfectly competitive firm faces a horizontal demand curve at the market price. We saw an example of a horizontal demand curve in the module on elasticity. Such a curve is perfectly elastic, meaning that any quantity is demanded at a given price.

Note that Figure 9.1 shows the market (and demand curve) for a **perfectly competitive industry** and Figure 9.3 shows the demand curve for a **perfectly competitive firm**.

This video demonstrates how average revenue equals marginal revenue, which equals price in a perfectly competitive market.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=199#oembed-2

Self Check: Perfectly Competitive Firms and Industries

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/652

173. Outcome: Costs and Revenue in a Perfectly Competitive Market

What you'll learn to do: calculate and graph the firm's fixed, variable, average, marginal, and total costs.

In this outcome, we learn how perfectly competitive firms make their one important decision of how much to produce based on the firm's costs.

The specific things you'll learn in this section include:

- Calculate and graph the firm's average, marginal and total revenues
- Determine the profit maximizing output level and price using graphs and demand schedules; is able to calculate and graphically illustrate where marginal revenue equals marginal costs

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: How Perfectly Competitive Firms Make Output Decisions
- Self Check: Costs and Revenues in a Perfectly Competitive Market

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

174. Reading: How Perfectly Competitive Firms Make Output Decisions

How Perfectly Competitive Firms Make Output Decisions

A ***perfectly competitive firm*** has only one major decision to make—namely, what quantity to produce. To understand why this is so, consider a different way of writing out the basic definition of ***profit***:

$$\begin{aligned} \text{Profit} &= \text{Total Revenue} - \text{Total Cost} \\ &= (\text{Price})(\text{Quantity Produced}) - (\text{Average Cost})(\text{Quantity Produced}) \end{aligned}$$

Since a perfectly competitive firm must accept the price for its output as determined by the product's market demand and supply, it cannot choose the price it charges. This is already determined in the profit equation, and so the perfectly competitive firm can sell any number of units at exactly the same price. It implies that the firm faces a perfectly elastic demand curve for its product: buyers are willing to buy any number of units of output from the firm at the market price. When the perfectly competitive firm chooses what quantity to produce, then this quantity—along with the prices prevailing in the market for output and inputs—will determine the firm's total revenue, total costs, and ultimately, level of profits.

DETERMINING THE HIGHEST PROFIT BY

COMPARING TOTAL REVENUE AND TOTAL COST

A perfectly competitive firm can sell as large a quantity as it wishes, as long as it accepts the prevailing market price. Total revenue is going to increase as the firm sells more, depending on the price of the product and the number of units sold. If you increase the number of units sold at a given price, then total revenue will increase. If the price of the product increases for every unit sold, then total revenue also increases. As an example of how a perfectly competitive firm decides what quantity to produce, consider the case of a small farmer who produces raspberries and sells them frozen for $4 per pack. Sales of one pack of raspberries will bring in $4, two packs will be $8, three packs will be $12, and so on. If, for example, the price of frozen raspberries doubles to $8 per pack, then sales of one pack of raspberries will be $8, two packs will be $16, three packs will be $24, and so on.

Total revenue and ***total costs*** for the raspberry farm, broken down into fixed and variable costs, are shown in Table 8.1 and also appear in Figure 8.2. The horizontal axis shows the quantity of frozen raspberries produced in packs; the vertical axis shows both total revenue and total costs, measured in dollars. The total cost curve intersects with the vertical axis at a value that shows the level of fixed costs, and then slopes upward. All these cost curves follow the same characteristics as the curves covered in the Cost and Industry Structure module.

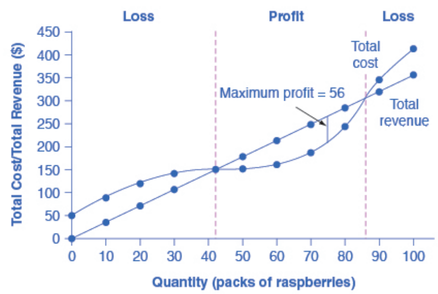


Figure 8.2. *Total Cost and Total Revenue at the Raspberry Farm. Total revenue for a perfectly competitive firm is a straight line sloping up. The slope is equal to the price of the good. Total cost also slopes up, but with some curvature. At higher levels of output, total cost begins to slope upward more steeply because of diminishing marginal returns. The maximum profit will occur at the quantity where the gap of total revenue over total cost is largest.*

Table 8.1 Total Cost and Total Revenue at the Raspberry Farm

Quantity (Q)	Total Cost (TC)	Fixed Cost (FC)	Variable Cost (VC)	Total Revenue (TR)	Profit
0	$62	$62	–	$0	−$62
10	$90	$62	$28	$40	−$50
20	$110	$62	$48	$80	−$30
30	$126	$62	$64	$120	−$6
40	$144	$62	$82	$160	$16
50	$166	$62	$104	$200	$34
60	$192	$62	$130	$240	$48
70	$224	$62	$162	$280	$56
80	$264	$62	$202	$320	$56
90	$324	$62	$262	$360	$36
100	$404	$62	$342	$400	−$4

Based on its total revenue and total cost curves, a perfectly competitive firm like the raspberry farm can calculate the quantity of output that will provide the highest level of profit. At any given quantity, total revenue minus total cost will equal profit. One way to determine the most profitable quantity to produce is to see at what quantity total revenue exceeds total cost by the largest amount. On Figure 8.2, the vertical gap between total revenue and total cost represents either profit (if total revenues are greater that total costs at a certain quantity) or losses (if total costs are greater that total revenues at a certain quantity). In this example, total costs will exceed total revenues at output levels from 0 to 40, and so over this range of output, the firm will be making losses. At output levels from 50 to 80, total revenues exceed total costs, so the firm is earning profits. But then at an output of 90 or 100, total costs again exceed total revenues and the firm is making losses. Total profits appear in the final column of Table 8.1. The highest total profits in the table, as in the figure that is based on the table values, occur at an output of 70–80, when profits will be $56.

A higher price would mean that total revenue would be higher for every quantity sold. A lower price would mean that total revenue would be lower for every quantity sold. What happens if the price drops low enough so that the total revenue line is completely below the total cost curve; that is, at every level of output, total costs are higher than total revenues? In this instance, the best the firm can do is to suffer losses. But a profit-maximizing firm will prefer the quantity of output where total revenues come closest to total costs and thus where the losses are smallest.

(Later we will see that sometimes it will make sense for the firm to shutdown, rather than stay in operation producing output.)

Watch the following video to learn more about the point of profit maximization.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=201#oembed-1

COMPARING MARGINAL REVENUE AND MARGINAL COSTS

Firms often do not have the necessary data they need to draw a complete total cost curve for all levels of production. They cannot be sure of what total costs would look like if they, say, doubled production or cut production in half, because they have not tried it. Instead, firms experiment. They produce a slightly greater or lower quantity and observe how profits are affected. In economic terms, this practical approach to maximizing profits means looking at how changes in production affect marginal revenue and marginal cost.

Figure 8.3 presents the marginal revenue and marginal cost curves based on the total revenue and total cost in Table 8.1. The ***marginal revenue*** curve shows the additional revenue gained from selling one more unit. As mentioned before, a firm in perfect competition faces a perfectly elastic demand curve for its product—that is, the firm's demand curve is a horizontal line drawn at the market price level. This also means that the firm's marginal revenue curve is the same as the firm's demand curve: Every time a consumer demands one more unit, the firm sells one more unit and revenue goes up by exactly the same amount equal to the market price. In this example, every time a pack of frozen raspberries is sold, the firm's revenue increases by $4. Table 8.2 shows an example of this. This condition only holds for price taking firms in perfect competition where: marginal revenue = price.

The formula for marginal revenue is:

$$\text{marginal revenue} = \frac{\text{change in total revenue}}{\text{change in quantity}}$$

Table 8.2 Marginal Revenue

Price	Quantity	Total Revenue	Marginal Revenue
$4	1	$4	–
$4	2	$8	$4
$4	3	$12	$4
$4	4	$16	$4

Notice that marginal revenue does not change as the firm produces more output. That is because the price is determined by supply and demand and does not change as the farmer produces more (keeping in mind that, due to the relative small size of each firm, increasing their supply has no impact on the total market supply where price is determined).

Since a perfectly competitive firm is a price taker, it can sell whatever quantity it wishes at the market-determined price. Marginal cost, the cost per additional unit sold, is calculated by dividing the change in total cost by the change in quantity. The formula for marginal cost is:

$$\text{marginal cost} = \frac{\text{change in total cost}}{\text{change in quantity}}$$

Ordinarily, marginal cost changes as the firm produces a greater quantity.

In the raspberry farm example, shown in Figure 8.3, Figure 8.4 and Table 8.3, marginal cost at first declines as production increases from 10 to 20 to 30 packs of raspberries—which represents the area of increasing marginal returns that is not uncommon at low levels of production. But then marginal costs start to increase, displaying the typical pattern of diminishing marginal returns. If the firm is producing at a quantity where MR > MC, like 40 or 50 packs of raspberries, then it can increase profit by increasing output because the marginal revenue is exceeding the marginal cost. If the firm is producing at a quantity where MC > MR, like 90 or 100 packs, then it can increase profit by reducing output because the reductions in marginal cost will exceed the reductions in marginal revenue. The firm's profit-maximizing choice of output will occur where MR = MC (or at a choice close to that point). You will notice that what occurs on the production side is exemplified on the cost side. This is referred to as duality.

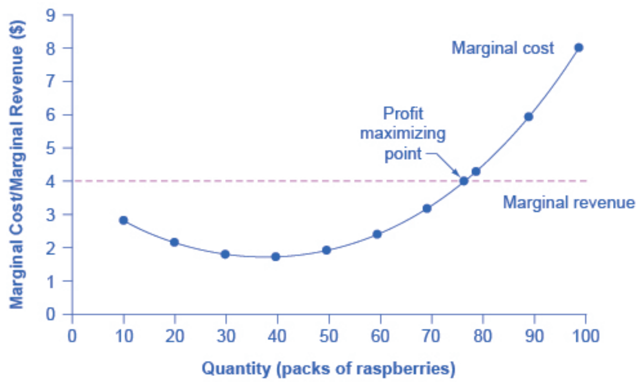


Figure 8.3. *Marginal Revenues and Marginal Costs at the Raspberry Farm: Individual Farmer. For a perfectly competitive firm, the marginal revenue (MR) curve is a horizontal straight line because it is equal to the price of the good, which is determined by the market, shown in Figure 8.4. The marginal cost (MC) curve is sometimes first downward-sloping, if there is a region of increasing marginal returns at low levels of output, but is eventually upward-sloping at higher levels of output as diminishing marginal returns kick in.*

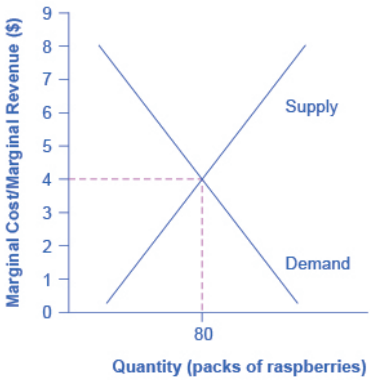


Figure 8.4. *Marginal Revenues and Marginal Costs at the Raspberry Farm: Raspberry Market. The equilibrium price of raspberries is determined through the interaction of market supply and market demand at $4.00.*

Table 8.3 Marginal Revenues and Marginal Costs at the Raspberry Farm

Quantity	Total Cost	Fixed Cost	Variable Cost	Marginal Cost	Total Revenue	Marginal Revenue
0	$62	$62	–	–	–	–
10	$90	$62	$28	$2.80	$40	$4.00
20	$110	$62	$48	$2.00	$80	$4.00
30	$126	$62	$64	$1.60	$120	$4.00
40	$144	$62	$82	$1.80	$160	$4.00
50	$166	$62	$104	$2.20	$200	$4.00
60	$192	$62	$130	$2.60	$240	$4.00
70	$224	$62	$162	$3.20	$280	$4.00
80	$264	$62	$202	$4.00	$320	$4.00
90	$324	$62	$262	$6.00	$360	$4.00
100	$404	$62	$342	$8.00	$400	$4.00

In this example, the marginal revenue and ***marginal cost*** curves cross at a price of $4 and a quantity of 80 produced. If the farmer started out producing at a level of 60, and then experimented with increasing production to 70, marginal revenues from the increase in production would exceed marginal costs—and so profits would rise. The farmer has an incentive to keep producing. From a level of 70 to 80, marginal cost and marginal revenue are equal so profit doesn't change. If the farmer then experimented further with increasing production from 80 to 90, he would find that marginal costs from the increase in production are greater than marginal revenues, and so profits would decline.

The profit-maximizing choice for a perfectly competitive firm will occur where marginal revenue is equal to marginal cost—that is, where MR = MC. A profit-seeking firm should keep expanding production as long as MR > MC. But at the level of output where MR = MC, the firm should recognize that it has achieved the highest possible level of economic profits. (In the example above, the profit maximizing output level is between 70 and 80 units of output, but the firm will not know they've maximized profit until they reach 80, where MR = MC.) Expanding production into the zone where MR < MC will only reduce economic profits. Because the marginal revenue received by a perfectly competitive firm is equal to the price P, so that P = MR, the profit-maximizing rule for a perfectly competitive firm can also be written as a recommendation to produce at the quantity where P = MC.

Self Check: Costs and Revenues in Competitive Markets

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/767

175. Outcome: Profit and Losses in a Perfectly Competitive Market

What you'll learn to do: measure variable and total costs as the area under the average variable and average total cost curves

In this outcome, you will analyze how perfectly competitive firms make output decisions about how much to produce. You'll look closer at a firm's cost curves to determine if a firm can afford to stay in business when prices are determined for them.

The specific things you'll learn in this section include:

- Measure total revenues as the area under the average revenue curves
- Calculate and graphically illustrate profit and losses for a perfectly competitive firm

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Profits and Losses with the Average Cost Curve
- Simulation: Maximizing Profit
- Self Check: Profit and Losses in Competitive Markets

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

176. Reading: Profits and Losses with the Average Cost Curve

Profits and Losses with the Average Cost Curve

Does maximizing profit (producing where MR = MC) imply an actual economic profit? The answer depends on the relationship between price and average total cost. If the price that a firm charges is higher than its average cost of production for that quantity produced, then the firm will earn profits. Conversely, if the price that a firm charges is lower than its average cost of production, the firm will suffer losses. You might think that, in this situation, the farmer may want to shut down immediately. Remember, however, that the firm has already paid for fixed costs, such as equipment, so it may continue to produce and incur a loss. Figure 8.5 illustrates three situations: (a) where price intersects marginal cost at a level above the average cost curve, (b) where price intersects marginal cost at a level equal to the average cost curve, and (c) where price intersects marginal cost at a level below the average cost curve.

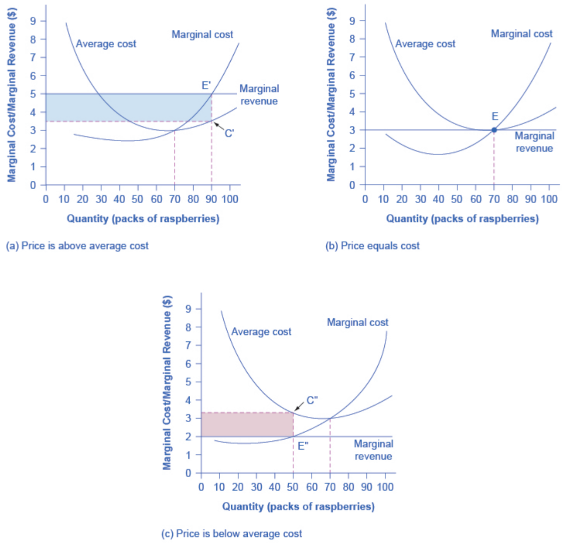


Figure 8.5. *Price and Average Cost at the Raspberry Farm In (a), price intersects marginal cost above the average cost curve. Since price is greater than average cost, the firm is making a profit. In (b), price intersects marginal cost at the minimum point of the average cost curve. Since price is equal to average cost, the firm is breaking even. In (c), price intersects marginal cost below the average cost curve. Since price is less than average cost, the firm is making a loss.*

First consider a situation where the price is equal to $5 for a pack of frozen raspberries. The rule for a profit-maximizing perfectly competitive firm is to produce the level of output where Price = MR = MC, so the raspberry farmer will produce a quantity of 90, which is labeled as E in Figure 8.5 (a). Remember that the area of a rectangle is equal to its base multiplied by its height. The farm's total revenue at this price will be shown by the larger rectangle starting from the origin right to a quantity of 90 packs (the base) up to point E (the height), left to the price of $5, and back down to the origin. The average cost of producing 90 packs is shown by point C or about $3.50. Total costs will be the quantity of 90 times the average cost of $3.50, which is shown by the area of the rectangle from the origin to a quantity of 90, up to point C, over to the vertical axis and down to the origin. It should be clear from examining the two rectangles that total revenue is greater than total cost. Thus, profits will be the blue shaded rectangle on top.

It can be calculated as:

profit	=	total revenue − total cost
	=	(90)($5.00) − (90)($3.50)
	=	$135

Or, it can be calculated as:

profit	=	(price − average cost) × quantity
	=	($5.00 − $3.50) × 90
	=	135

Now consider Figure 8.5 (b), where the price has fallen to $3.00 for a pack of frozen raspberries. Again, the perfectly competitive firm will choose the level of output where Price = MR = MC, but in this case, the quantity produced will be 70. At this price and output level, where the marginal cost curve is crossing the average cost curve, the price received by the firm is exactly equal to its average cost of production.

The farm's total revenue at this price will be shown by the large rectangle from the origin over to a quantity of 70 packs (the base) up to point E' (the height), over to the price of $3, and back to the origin. The average cost of producing 70 packs is the same as the farm's total revenue because the price of one unit is equal to the cost to produce that unit. Thus, the firm is making zero profit. The calculations are as follows:

profit	=	total revenue – total cost
	=	(70)($3.00) – (70)($3.00)
	=	$0

Or, it can be calculated as:

profit	=	(price – average cost) × quantity
	=	($3.00 – $3.00) × 70
	=	$0

In Figure 8.5 (c), the market price has fallen still further to $2.00 for a pack of frozen raspberries. At this price, marginal revenue intersects marginal cost at a quantity of 50. The farm's total revenue at this price will be shown by the rectangle from the origin over to a quantity of 50 packs (the base) up to point E" (the height), over to the price of $2, and back to the origin. The average cost of producing 50 packs is shown by point C" or about $3.30. Total costs will be the quantity of 50 times the average cost of $3.30, which is shown by the area of the rectangle from the origin to a quantity of 50, up to point C", over to the vertical axis and down to the origin. It should be clear from examining the two rectangles that total revenue is less than total cost. Thus, the firm is losing money and the loss (or negative profit) will be the rose-shaded rectangle.

The calculations are:

profit	=	total revenue − total cost
	=	(50)($2.00) − (50)($3.30)
	=	-$65.00

Or:

profit	=	(price − average cost) × quantity
	=	($1.75 − $3.30) × 50
	=	-$65.00

If the market price received by a perfectly competitive firm leads it to produce at a quantity where the price is greater than average cost, the firm will earn profits. If the price received by the firm causes it to produce at a quantity where price equals average cost, which occurs at the minimum point of the AC curve, then the firm earns zero profits. Finally, if the price received by the firm leads it to produce at a quantity where the price is less than average cost, the firm will earn losses. This is summarized in Table 8.4.

Table 8.4

If…	**Then…**
Price > ATC	Firm earns an economic profit
Price = ATC	Firm earns zero economic profit
Price < ATC	Firm earns a loss

Watch this video to see the concept of economic profit explained:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=203#oembed-1

Self Check: Profit and Losses in Competitive Markets

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/765

177. Simulation: Maximizing Profit

Try It

Play the simulation below multiple times to see how different choices lead to different outcomes. All simulations allow unlimited attempts so that you can gain experience applying the concepts.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=204

178. Outcome: The Shutdown Point

What you'll learn to do: determine the break-even, and the shutdown points of production for a perfectly competitive firm

In this outcome, you will see why it is sometimes advantageous for a firm to continue producing even when it is experiencing losses.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: The Shutdown Point
- Self Check: The Shutdown Point

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

179. Reading: The Shutdown Point

The Shutdown Point

The possibility that a firm may earn losses raises a question: Why can the firm not avoid losses by shutting down and not producing at all? The answer is that shutting down can reduce variable costs to zero, but in the short run, the firm has already paid for fixed costs. As a result, if the firm produces a quantity of zero, it would still make losses because it would still need to pay for its fixed costs. So, when a firm is experiencing losses, it must face a question: should it continue producing or should it shut down?

As an example, consider the situation of the Yoga Center, which has signed a contract to rent space that costs $10,000 per month. If the firm decides to operate, its marginal costs for hiring yoga teachers is $15,000 for the month. If the firm shuts down, it must still pay the rent, but it would not need to hire labor. Let's take a look at three possible scenarios. In the first scenario, the Yoga Center does not have any clients, and therefore does not make any revenues, in which case it faces losses of $10,000 equal to the fixed costs. In the second scenario, the Yoga Center has clients that earn the center revenues of $10,000 for the month, but ultimately experiences losses of $15,000 due to having to hire yoga instructors to cover the classes. In the third scenario, the Yoga Center earns revenues of $20,000 for the month, but experiences losses of $5,000.

In all three cases, the Yoga Center loses money. In all three cases, when the rental contract expires in the long run, assuming revenues do not improve, the firm should exit this business. In the short run, though, the decision varies depending on the level of losses and whether the firm can cover its variable costs. In scenario 1, the center does not have any revenues, so hiring yoga teachers would increase variable costs and losses, so it should shut down and only incur its fixed costs. In scenario 2, the center's losses are greater because it does not make enough revenue to offset the increased variable costs plus fixed costs, so it should shut down immediately. If price is below the minimum average variable cost, the firm must shut down. In contrast, in scenario 3 the revenue that the center can earn is high enough that the losses diminish when it remains open, so the center should remain open in the short run.

Should the Yoga Center Shut Down Now or Later?

Scenario 1

If the center shuts down now, revenues are zero but it will not incur any variable costs and would only need to pay fixed costs of $10,000.

profit = total revenue – (fixed costs + variable cost)

profit = 0 – $10,000 = –$10,000

Scenario 2

The center earns revenues of $10,000, and variable costs are $15,000. The center should shut down now.

profit = total revenue – (fixed costs + variable cost)

profit = $10,000 – ($10,000 + $15,000) = –$15,000

Scenario 3

The center earns revenues of $20,000, and variable costs are $15,000. The center should continue in business.

profit = total revenue − (fixed costs + variable cost)

profit = $20,000 − ($10,000 + $15,000) = −$5,000

This example suggests that the key factor is whether a firm can earn enough revenues to cover at least its variable costs by remaining open. Let's return now to our raspberry farm. Figure 8.6 illustrates this lesson by adding the average variable cost curve to the marginal cost and average cost curves. At a price of $2.20 per pack, as shown in Figure 8.6 (a), the farm produces at a level of 50. It is making losses of $56 (as explained earlier), but price is above average variable cost and so the firm continues to operate. However, if the price declined to $1.80 per pack, as shown in Figure 8.6 (b), and if the firm applied its rule of producing where P = MR = MC, it would produce a quantity of 40. This price is below average variable cost for this level of output. If the farmer cannot pay workers (the variable costs), then it has to shut down. At this price and output, total revenues would be $72 (quantity of 40 times price of $1.80) and total cost would be $144, for overall losses of $72. If the farm shuts down, it must pay only its fixed costs of $62, so shutting down is preferable to selling at a price of $1.80 per pack.

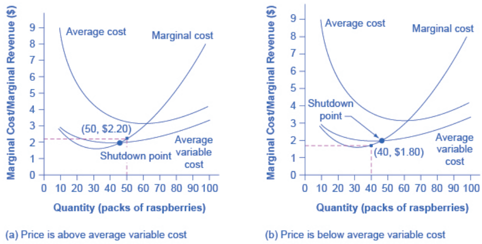


Figure 8.6. *The Shutdown Point for the Raspberry Farm. In (a), the farm produces at a level of 50. It is making losses of $56, but price is above average variable cost, so it continues to operate. In (b), total revenues are $72 and total cost is $144, for overall losses of $72. If the farm shuts down, it must pay only its fixed costs of $62. Shutting down is preferable to selling at a price of $1.80 per pack.*

Looking at Table 8.6, if the price falls below $2.05, the minimum average variable cost, the firm must shut down.

Table 8.6. Cost of Production for the Raspberry Farm

Quantity	Total Cost	Fixed Cost	Variable Cost	Marginal Cost	Average Cost	Average Variable Cost
0	$62	$62	–	–	–	–
10	$90	$62	$28	$2.80	$9.00	$2.80
20	$110	$62	$48	$2.00	$5.50	$2.40
30	$126	$62	$64	$1.60	$4.20	$2.13
40	$144	$62	$82	$1.80	$3.60	$2.05
50	$166	$62	$104	$2.20	$3.32	$2.08
60	$192	$62	$130	$2.60	$3.20	$2.16
70	$224	$62	$162	$3.20	$3.20	$2.31
80	$264	$62	$202	$4.00	$3.30	$2.52
90	$324	$62	$262	$6.00	$3.60	$2.91
100	$404	$62	$342	$8.00	$4.04	$3.42

The intersection of the average variable cost curve and the marginal cost curve, which shows the price where the firm would lack enough revenue to cover its variable costs, is called the **shutdown point**. If the perfectly competitive firm can charge a price above the shutdown point, then the firm is at least covering its average variable costs. It is also making enough revenue to cover at least a portion of fixed costs, so it should limp ahead even if it is making losses in the short run, since at least those losses will be smaller than if the firm shuts down immediately and incurs a loss equal to total fixed costs. However, if the firm is receiving a price below the price at the shutdown point, then the firm is not even covering its variable costs. In this case, staying open is making the firm's losses larger, and it should shut down immediately. To summarize, if:

- price < minimum average variable cost, then firm shuts down
- price = minimum average variable cost, then firm stays in business

SHORT-RUN OUTCOMES FOR PERFECTLY COMPETITIVE FIRMS

The average cost and average variable cost curves divide the marginal cost curve into three segments, as shown in Figure 8.7. At the market price, which the perfectly competitive firm accepts as given, the profit-maximizing firm chooses the output level where price or marginal revenue, which are the same thing for a perfectly competitive firm, is equal to marginal cost: P = MR = MC.

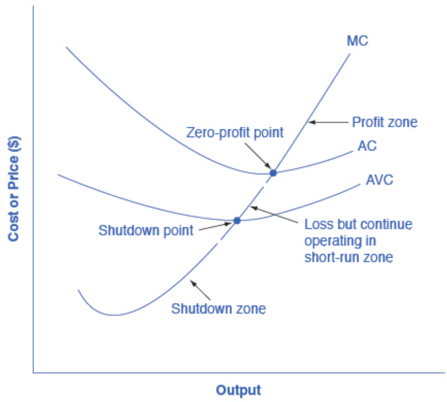


Figure 8.7. *Profit, Loss, Shutdown. The marginal cost curve can be divided into three zones, based on where it is crossed by the average cost and average variable cost curves. The point where MC crosses AC is called the zero-profit point. If the firm is operating at a level of output where the market price is at a level higher than the zero-profit point, then price will be greater than average cost and the firm is earning profits. If the price is exactly at the zero-profit point, then the firm is making zero profits. If price falls in the zone between the shutdown point and the*

zero-profit point, then the firm is making losses but will continue to operate in the short run, since it is covering its variable costs. However, if price falls below the price at the shutdown point, then the firm will shut down immediately, since it is not even covering its variable costs.

First consider the upper zone, where prices are above the level where marginal cost (MC) crosses average cost (AC) at the zero profit point. At any price above that level, the firm will earn profits in the short run. If the price falls exactly on the zero profit point where the MC and AC curves cross, then the firm earns zero profits. If a price falls into the zone between the zero profit point, where MC crosses AC, and the shutdown point, where MC crosses AVC, the firm will be making losses in the short run—but since the firm is more than covering its variable costs, the losses are smaller than if the firm shut down immediately. Finally, consider a price at or below the shutdown point where MC crosses AVC. At any price like this one, the firm will shut down immediately, because it cannot even cover its variable costs.

Watch this video to see an illustrated example of zero profit, or the normal profit, point:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=206#oembed-1

MARGINAL COST AND THE FIRM'S SUPPLY CURVE

For a perfectly competitive firm, the marginal cost curve is identical to the firm's supply curve starting from the minimum point on the average variable cost curve. To understand why this perhaps surprising insight holds true, first think about what the supply curve means. A firm checks the market price and then looks at its supply curve to decide what quantity to produce. Now, think about what it means to say that a firm will maximize its profits by producing at the quantity where P = MC. This rule means that the firm checks the market price, and then looks at its marginal cost to determine the quantity to produce—and makes sure that the price is greater than the minimum average variable cost. In other words, the marginal cost curve above the minimum point on the average variable cost curve becomes the firm's supply curve.

LINK IT UP

Watch this video that addresses how drought in the United States can impact food prices across the world. (Note that the story on the drought is the second one in the news report; you need to let the video play through the first story in order to watch the story on the drought.)

As discussed in the module on Demand and Supply, many of the reasons that supply curves shift relate to underlying changes in costs. For example, a lower price of key inputs or new technologies that reduce production costs cause supply to shift to the right; in contrast, bad weather or added government regulations can add to costs of certain goods in a way that causes supply to shift to the left. These shifts in the firm's supply curve can also be interpreted as shifts of the marginal cost curve. A shift in costs of production that increases marginal costs at all levels of output—and shifts MC to the left—will cause a perfectly competitive firm to produce less at any given market price. Conversely, a shift in costs of production that decreases marginal costs at all levels of output will shift MC to the right and as a result, a competitive firm will choose to expand its level of output at any given price.

AT WHAT PRICE SHOULD THE FIRM CONTINUE PRODUCING IN THE SHORT RUN?

To determine the short-run economic condition of a firm in perfect competition, follow the steps outlined below. Use the data shown in Table 8.7 below:

Table 8.7 Calculating Short-Run Economic Condition

Q	P	TFC	TVC	TC	AVC	ATC	MC	TR	Profits
0	$28	$20	$0	–	–	–	–	–	–
1	$28	$20	$20	–	–	–	–	–	–
2	$28	$20	$25	–	–	–	–	–	–
3	$28	$20	$35	–	–	–	–	–	–
4	$28	$20	$52	–	–	–	–	–	–
5	$28	$20	$80	–	–	–	–	–	–

Step 1. Determine the cost structure for the firm. For a given total fixed costs and variable costs, calculate total cost, average variable cost, average total cost, and marginal cost. Follow the formulas given in the Cost and Industry Structure module. These calculations are shown in Table 8.8 below:

Table 8.8

Q	P	TFC	TVC	TC (TFC+TVC)	AVC (TVC/Q)	ATC (TC/Q)	MC $(TC_2-TC_1)/(Q_2-Q_1)$
0	$28	$20	$0	$20+$0=$20	–	–	–
1	$28	$20	$20	$20+$20=$40	$20/1=$20.00	$40/1=$40.00	($40–$20)/(1–0)= $20
2	$28	$20	$25	$20+$25=$45	$25/2=$12.50	$45/2=$22.50	($45–$40)/(2–1)= $5
3	$28	$20	$35	$20+$35=$55	$35/3=$11.67	$55/3=$18.33	($55–$45)/(3–2)= $10
4	$28	$20	$52	$20+$52=$72	$52/4=$13.00	$72/4=$18.00	($72–$55)/(4–3)= $17
5	$28	$20	$80	$20+$80=$100	$80/5=$16.00	$100/5=$20.00	($100–$72)/(5–4)= $28

Step 2. Determine the market price that the firm receives for its product. This should be given information, as the firm in perfect competition is a price taker. With the given price, calculate total revenue as equal to price multiplied by quantity for all output levels produced. In this example, the given price is $30. You can see that in the second column of Table 8.9.

Table 8.9

Quantity	Total Revenue	Total C
0	$0	$20
1	$28	$40
2	$56	$45
3	$84	$55
4	$112	$72
5	$140	$100

Step 4. To find the profit-maximizing output level, look at the Marginal Cost column (at every output level produced), as shown in Table 8.11, and determine where it is equal to the market price. The output level where price equals the marginal cost is the output level that maximizes profits.

Table 8.11

Q	P	TFC	TVC	TC	AVC	ATC	MC	TR	Profits
0	$28	$20	$0	$20	–	–	–	$0	−$20
1	$28	$20	$20	$40	$20.00	$40.00	$20	$28	−$12
2	$28	$20	$25	$45	$12.50	$22.50	$5	$56	$11
3	$28	$20	$35	$55	$11.67	$18.33	$10	$84	$29
4	$28	$20	$52	$72	$13.00	$18.00	$17	$112	$40
5	**$28**	$20	$80	$100	$16.40	$20.40	**$30**	$140	**$40**

Step 5. Once you have determined the profit-maximizing output level (in this case, output quantity 5), you can look at the amount of profits made (in this case, $50).

Step 6. If the firm is making economic losses, the firm needs to determine whether it produces the output level where price equals marginal revenue and equals marginal cost or it shuts down and only incurs its fixed costs.

Step 7. For the output level where marginal revenue is equal to marginal cost, check if the market price is greater than the average variable cost of producing that output level.

- If P > AVC but P < ATC, then the firm continues to produce in the short-run, making economic losses.
- If P < AVC, then the firm stops producing and only incurs its fixed costs.

In this example, the price of $30 is greater than the AVC ($16.40) of producing 5 units of output, so the firm continues producing.

Watch this video to see an illustrated example of a firm who is facing loses:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=206#oembed-2

Self Check: The Shutdown Point

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether

to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/659

180. Outcome: Entry and Exit Decisions

What you'll learn to do: explain the difference between short-run and long-run equilibrium

This section will focus on the long-run adjustment process. In the short run, firms do not have control over their fixed inputs, but in the long run, they can adjust their factors of production to remain productive.

The specific things you'll learn to do in this section include:

- Explain the concept of "zero economic profit"

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Entry and Exit Decisions in the Long Run
- Self Check: Entry and Exit Decisions and Zero Economic Profit

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

181. Reading: Entry and Exit Decisions in the Long Run

Entry and Exit Decisions in the Long Run

The line between the short run and the long run cannot be defined precisely with a stopwatch, or even with a calendar. It varies according to the specific business. The distinction between the short run and the long run is therefore more technical: in the short run, firms cannot change the usage of fixed inputs, while in the long run, the firm can adjust all factors of production.

In a competitive market, profits are a red cape that incites businesses to charge. If a business is making a profit in the short run, it has an incentive to expand existing factories or to build new ones. New firms may start production, as well. When new firms enter the industry in response to increased industry profits it is called ***entry***.

Losses are the black thundercloud that causes businesses to flee. If a business is making losses in the short run, it will either keep limping along or just shut down, depending on whether its revenues are covering its variable costs. But in the long run, firms that are facing losses will shut down at least some of their output, and some firms will cease production altogether. The long-run process of reducing production in response to a sustained pattern of losses is called ***exit***.

Why do firms cease to exist?

Can we say anything about what causes a firm to exit an industry?

Profits are the measurement that determines whether a business stays operating or not. Individuals start businesses with the purpose of making profits. They invest their money, time, effort, and many other resources to produce and sell something that they hope will give them something in return. Unfortunately, not all businesses are successful, and many new startups soon realize that their "business adventure" must eventually end.

In the model of perfectly competitive firms, those that consistently cannot make money will "exit," which is a nice, bloodless word for a more painful process. When a business fails, after all, workers lose their jobs, investors lose their money, and owners and managers can lose their dreams. Many businesses fail. The U.S. Small Business Administration indicates that in 2009–2010, for example, 533,945 firms "entered" in the United States, but 593,347 firms "exited." About 96.3% and 96.6% of these business entries and exits, respectively, involved small firms with fewer than 20 employees.

Sometimes a business fails because of poor management or workers who are not very productive, or because of tough domestic or foreign competition. Businesses also fail from a variety of causes that might best be summarized as bad luck. For example, conditions of demand and supply in the market shift in an unexpected way, so that the prices that can be charged for outputs fall or the prices that need to be paid for inputs rise. With millions of businesses in the U.S. economy, even a small fraction of them failing will affect many people—and business failures can be very hard on the workers and managers directly involved. But from the standpoint of the overall economic system, business exits are sometimes a necessary evil if a market-oriented system is going to offer a flexible mechanism for satisfying customers, keeping costs low, and inventing new products.

How Entry and Exit Lead to Zero Profits in the

Long Run

No ***perfectly competitive firm*** acting alone can affect the market price. However, the combination of many firms entering or exiting the market will affect overall supply in the market. In turn, a shift in supply for the market as a whole will affect the market price. Entry and exit to and from the market are the driving forces behind a process that, in the long run, pushes the price down to minimum average total costs so that all firms are earning a zero profit.

To understand how short-run profits for a perfectly competitive firm will evaporate in the long run, imagine the following situation. The market is in ***long-run equilibrium***, where all firms earn zero economic profits producing the output level where P = MR = MC and P = AC. No firm has the incentive to enter or leave the market. Let's say that the product's demand increases, and with that, the market price goes up. The existing firms in the industry are now facing a higher price than before, so they will increase production to the new output level where P = MR = MC.

This will temporarily make the market price rise above the average cost curve, and therefore, the existing firms in the market will now be earning economic profits. However, these economic profits attract other firms to enter the market. Entry of many new firms causes the market supply curve to shift to the right. As the supply curve shifts to the right, the market price starts decreasing, and with that, economic profits fall for new and existing firms. As long as there are still profits in the market, entry will continue to shift supply to the right. This will stop whenever the market price is driven down to the zero-profit level, where no firm is earning economic profits.

Short-run losses will fade away by reversing this process. Say that the market is in long-run equilibrium. This time, instead, demand decreases, and with that, the market price starts falling. The existing firms in the industry are now facing a lower price than before, and as it will be below the average cost curve, they will now be making economic losses. Some firms will continue producing where the new P = MR = MC, as long as they are able to cover their average variable costs. Some firms will have to shut down immediately as they will not be able to cover their average variable costs, and will then only incur their fixed costs, minimizing their losses. Exit of many firms causes the market supply curve to shift to the left. As the supply curve shifts to the left, the market price starts rising, and economic losses start to be lower. This process ends whenever the market price rises to the zero-profit level, where the existing firms are no longer losing money and are at zero profits again. Thus, while a perfectly competitive firm can earn profits in the short run, in the long run the process of entry will push down prices until they reach the zero-profit level. Conversely, while a perfectly competitive firm may earn losses in the short run, firms will not continually lose money. In the long run, firms making losses are able to escape from their fixed costs, and their exit from the market will push the price back up to the zero-profit level. In the long run, this process of entry and exit will drive the price in perfectly competitive markets to the zero-profit point at the bottom of the AC curve, where marginal cost crosses average cost.

The Long-Run Adjustment and Industry Types

Whenever there are expansions in an industry, costs of production for the existing and new firms could either stay the same, increase, or even decrease. Therefore, we can categorize an industry as being (1) a constant cost industry (as demand increases, the cost of production for firms stays the same), (2) an increasing cost industry (as demand increases, the cost of production for firms increases), or (3) a decreasing cost industry (as demand increases the costs of production for the firms decreases).

For a ***constant cost industry***, whenever there is an increase in market demand and price, then the supply curve shifts to the right with new firms' entry and stops at the point where the new long-run equilibrium intersects at the same market price as before. But why will costs remain the same? In this type of industry, the supply curve is very elastic. Firms can easily supply any quantity that consumers demand. In addition, there is a perfectly elastic supply of inputs—firms can easily increase their demand for employees, for example, with no increase to wages. As discussed earlier, increased demand for ethanol in recent years has caused the demand for corn to increase. Consequently, many farmers switched from growing wheat to growing corn. Agricultural markets are generally good examples of constant cost industries.

For an ***increasing cost industry***, as the market expands, the old and new firms experience increases in their costs of production, which makes the new zero-profit level intersect at a higher price than before. Here companies may have to deal with limited inputs, such as skilled labor. As the demand for these workers rise, wages rise and this increases the cost of production for all firms. The industry supply curve in this type of industry is more inelastic.

For a ***decreasing cost industry***, as the market expands, the old and new firms experience lower costs of production, which makes the new zero-profit level intersect at a lower price than before. In this case, the industry and all the firms in it are experiencing falling average total costs. This can be due to an improvement in technology in the entire industry or an increase in the education of employees. High tech industries may be a good example of a decreasing cost market.

Figure 8.8 (a) presents the case of an adjustment process in a constant cost industry. Whenever there are output expansions in this type of industry, the long-run outcome implies more output produced at exactly the same original price. Note that supply was able to increase to meet the increased demand. When we join the before and after long-run equilibriums, the resulting line is the long run supply (LRS) curve in perfectly competitive markets. In this case, it is a flat curve. Figure 8.8 (b) and Figure 8.8 (c) present the cases for an increasing cost and decreasing cost industry, respectively. For an increasing cost industry, the LRS is upward sloping, while for a decreasing cost industry, the LRS is downward sloping.

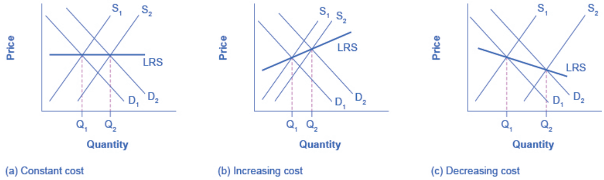


Figure 8.8. *Adjustment Process in a Constant-Cost Industry. In (a), demand increased and supply met it. Notice that the supply increase is equal to the demand increase. The result is that the equilibrium price stays the same as quantity sold increases. In (b), notice that sellers were not able to increase supply as much as demand. Some inputs were scarce, or wages were rising. The equilibrium price rises. In (c), sellers easily increased supply in response to the demand increase. Here, new technology or economies of scale caused the large increase in*

supply, resulting in declining equilibrium price.

Self Check: Entry and Exit Decisions and Zero Economic Profit

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/768

182. Outcome: Efficiency in Perfectly Competitive Markets

What you'll learn to do: understand why perfectly competitive markets are efficient

In this outcome, you'll see how perfectly competitive firms can be both allocatively and productively efficient. This means these markets allocate resources without wasting and also provide the maximal amount of satisfaction for the society.

LEARNING ACTIVITIES

The learning activities for this section include:

- Reading: Efficiency in Perfectly Competitive Markets
- Self Check: Perfectly Competitive Markets and Efficiency

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

183. Reading: Efficiency in Perfectly Competitive Markets

Efficiency in Perfectly Competitive Markets

When profit-maximizing firms in perfectly competitive markets combine with utility-maximizing consumers, something remarkable happens: the resulting quantities of outputs of goods and services demonstrate both productive and allocative efficiency (terms that were first introduced in the Choice in a World of Scarcity section of the Introduction to Economics and Scarcity module).

Productive efficiency means producing without waste, so that the choice is on the production possibility frontier. In the long run in a perfectly competitive market, because of the process of entry and exit, the price in the market is equal to the minimum of the long-run average cost curve. In other words, goods are being produced and sold at the lowest possible average cost.

Allocative efficiency means that among the points on the production possibility frontier, the point that is chosen is socially preferred—at least in a particular and specific sense. In a perfectly competitive market, price will be equal to the marginal cost of production. Think about the price that is paid for a good as a measure of the social benefit received for that good; after all, willingness to pay conveys what the good is worth to a buyer. Then think about the marginal cost of producing the good as representing not just the cost for the firm, but more broadly as the social cost of producing that good. When perfectly competitive firms follow the rule that profits are maximized by producing at the quantity where price is equal to marginal cost, they are thus ensuring that the social benefits received from producing a good are in line with the social costs of production.

To explore what is meant by allocative efficiency, it is useful to walk through an example. Begin by assuming that the market for wholesale flowers is perfectly competitive, and so P = MC. Now, consider what it would mean if firms in that market produced a lesser quantity of flowers. At a lesser quantity, marginal costs will not yet have increased as much, so that price will exceed marginal cost; that is, P > MC. In that situation, the benefit to society as a whole of producing additional goods, as measured by the willingness of consumers to pay for marginal units of a good, would be higher than the cost of the inputs of labor and physical capital needed to produce the marginal good. In other words, the gains to society as a whole from producing additional marginal units will be greater than the costs.

Conversely, consider what it would mean if, compared to the level of output at the allocatively efficient choice when P = MC, firms produced a greater quantity of flowers. At a greater quantity, marginal costs of production will have increased so that P < MC. In that case, the marginal costs of producing additional flowers is greater than the benefit to society as measured by what people are willing to pay. For society as a whole, since the costs are outstripping the benefits, it will make sense to produce a lower quantity of such goods.

When perfectly competitive firms maximize their profits by producing the quantity where P = MC, they also assure that the benefits to consumers of what they are buying, as measured by the price they are willing to pay, is equal to the costs to society of producing the marginal units, as measured by the marginal costs the firm must pay—and thus that allocative efficiency holds.

The statements that a perfectly competitive market in the long run will feature both productive and allocative efficiency do need to be taken with a few grains of salt. Remember, economists are using the concept of "efficiency" in a particular and specific sense, not as a synonym for "desirable in every way." For one thing, consumers' ability to pay reflects the income distribution in a particular society. Thus, a homeless person may have no ability to pay for housing because they have insufficient income.

Perfect competition, in the long run, is a hypothetical benchmark. For market structures such as monopoly, monopolistic competition, and oligopoly, which are more frequently observed in the real world than perfect competition, firms will not always produce at the minimum of average cost, nor will they always set price equal to marginal cost. Thus, these other competitive situations will not produce productive and allocative efficiency.

Moreover, real-world markets include many issues that are assumed away in the model of perfect competition, including pollution, inventions of new technology, poverty which may make some people unable to pay for basic necessities of life, government programs like national defense or education, discrimination in labor markets, and buyers and sellers who must deal with imperfect and unclear information. These issues are explored in other modules. However, the theoretical efficiency of perfect competition does provide a useful benchmark for comparing the issues that arise from these real-world problems.

LINK IT UP

A quick glance at the Table below reveals the dramatic increase in North Dakota corn production—more than double. Taking into consideration that corn typically yields two to three times as many bushels per acre as wheat, it is obvious there has been a significant increase in bushels of corn. Why the increase in corn acreage? Converging prices.

Year	Corn (thousands of acres)	Wheat (thousands of acres)
1997	780	11,625
2006	1,690	8,800
(Source: USDA National Agricultural Statistics Service)		

Historically, wheat prices have been higher than corn prices, offsetting wheat's lower yield per acre. However, in recent years wheat and corn prices have been converging. In April 2013, A*gweek* reported the gap was just 71 cents per bushel. As the difference in price narrowed, switching to the production of higher yield per acre of corn simply made good business sense. Erik Younggren, president of the National Association of Wheat Growers said in the A*gweek* article, "I don't think we're going to see mile after mile of waving amber fields [of wheat] anymore." (Until wheat prices rise, we will probably be seeing field after field of tasseled corn.)

Key Concepts and Summary

Long-run equilibrium in perfectly competitive markets meets two important conditions: allocative efficiency and productive efficiency. These two conditions have important implications. First, resources are allocated to their best alternative use. Second, they provide the maximum satisfaction attainable by society.

Self Check: Perfectly Competitive Markets and Efficiency

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/766

184. Putting It Together: Perfect Competition

Summary

The goal of this module was analyze a firm's profit maximizing decisions under conditions of perfect competition. You learned how to:

- Define the characteristics of Perfect Competition
- Understand the difference between the firm and the industry
- Calculate and graph the firm's fixed, variable, average, marginal and total costs
- Measure variable and total costs as the area under the average variable and average total cost curves
- Determine the break-even, and the shutdown points of production for a perfectly competitive firm
- Explain the difference between short run and long run equilibrium
- Understand why perfectly competitive markets are efficient

Examples

Healthy Red Tomatoes by EpSos.de, CC-BY.

Let's return to the questions posed at the beginning of the module. Why do all tomatoes of the same type cost the same price in a farmer's market? The answer is because a farmer's market or a bunch of roadside tomato stands fit the characteristics of perfect competition: many firms (or sellers at the market), all selling a similar if not identical product, where it is easy for buyers and sellers to see what everyone is charging. In this situation, if one seller charged a higher price, none of the customers would do business with that seller, since they could get an identical product from another seller at the lower price. Why would a seller charge a lower price if they can sell all their inventory by the end of the day at the going price?

By contrast, why do different gas stations on the same strip of highway charge different amounts per gallon of gasoline? Why do different pizza restaurants charge different prices for the same product, say a large one topping pizza? We'll learn the answers to these questions in future modules.

Finally, what's so perfect about perfect competition? The answer is that perfect competition shows markets operating at their best. Perfect competition is productively efficient, because in the long run firms produce their products as cheaply as possible (i.e. at minimum average cost). What this means in a larger context is that the economy is operating on its production possibilities frontier, rather than inside the frontier (for more on the production possibilities frontier, click HERE).

It's also allocatively efficient, meaning it's producing the optimal quantity of output (i.e. where price equals marginal cost), because that quantity maximizes total economic surplus (for a review of surplus, click HERE). What this means is that it's producing at the right point on the production possibilities frontier.

Watch the following video for a summary and example of perfect competition:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=211#oembed-1

There is a lot going for the model of perfect competition. As a result, it's the model we typically use to compare real world industries against.

Let's turn now to those real world industries.

185. Glossary: Perfect Competition

entry
: the long-run process of firms entering an industry in response to industry profits

exit
: the long-run process of firms reducing production and shutting down in response to industry losses

long-run equilibrium
: where all firms earn zero economic profits producing the output level where P = MR = MC and P = AC

marginal revenue
: the additional revenue gained from selling one more unit

market structure
: the conditions in an industry, such as number of sellers, how easy or difficult it is for a new firm to enter, and the type of products that are sold

perfect competition
: each firm faces many competitors that sell identical products

price taker
: a firm in a perfectly competitive market that must take the prevailing market price as given

shutdown point
: level of output where the marginal cost curve intersects the average variable cost curve at the minimum point of AVC; if the price is below this point, the firm should shut down immediately

186. Discussion: Independent Trucking Analysis

Independent trucking is an industry that can be considered perfectly competitive. Draw a graph showing market supply, market demand, and equilibrium price and quantity. Draw a corresponding graph for the individual firm/trucker using the market equilibrium price and marginal cost curve. If you line up the two graphs horizontally, the equilibrium price should be the same on both graphs.

Now suppose that GDP increases as U.S. manufacturers produce more output. What impact will this have on the independent trucking industry in the short run, in terms of the market price, output of an individual firm, and market equilibrium quantity? Explain your reasoning. What impact will the increase in manufacturing output have in the long run? Show graphically and explain your reasoning.

PART XI

MODULE: MONOPOLY

187. Why It Matters: Monopoly

Why analyze a firm's profit maximizing strategies under conditions of a monopoly?

If perfect competition is at one end of the competitive spectrum, at the other end is monopoly. Mono means one. A monoplane is an aircraft with one wing. A monocle is a single eyeglass. Monopoly is a single supplier, the only firm in an industry. Monopolies have monopoly power, which is the ability to set the market price.

Monoplane by Robert Payne, CC-BY.

As you work through this module, think about the following questions:

- What prevents a monopoly from charging an infinite price?
- What is similar about the model of monopoly compared to perfect competition?
- What is different about the model of monopoly compared to perfect competition?

There are more industries which are monopolies than are perfectly competitive, but examples of pure monopoly are still hard to find in the U.S.. Google is not a monopoly. Nor is Microsoft or Amazon.com. Still we can learn a lot about how those firms operate by understanding the model of monopoly. Your local power company is a monopoly, but it doesn't operate exactly the way this module explains. How do we explain this anomaly? Let's find out.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=215

LEARNING OUTCOMES

- Define the characteristics of a monopoly
- Define and explain the sources of barriers to entry
- Calculate and graph a monopoly's fixed, variable, average, marginal and total costs
- Explain why a monopoly is inefficient using deadweight loss
- Analyze different strategies to control monopolies, including natural monopolies

188. Outcome: Introduction to Monopolies

What you'll learn to do: define the characteristics of a monopoly

Businesses love the idea of a free-market economy, so why then, would a business want to become a monopolist? Just think…if a business can control the entire market for a product, then they eliminate competition and are almost guaranteed a profit. In this outcome, you'll learn about what a monopoly is and why firms try to eliminate their competition.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Introduction to Monopolies
- Self Check: Introduction to Monopolies

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

189. Reading: Introduction to Monopolies

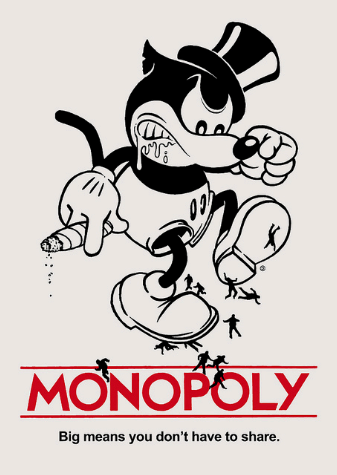
The Rest is History

Many of the opening case studies have focused on current events. This one steps into the past to observe how monopoly, or near monopolies, have helped shape history. In the spring of 1773, the East India Company, a firm that, in its time, was designated 'too big to fail,' was continuing to experience financial difficulties. To help shore up the failing firm, the British Parliament authorized the Tea Act. The act continued the tax on teas and made the East India Company the sole legal supplier of tea to the American colonies. By November, the citizens of Boston had had enough. They refused to permit the tea to be unloaded, citing their main complaint: "No taxation without representation." Arriving tea-bearing ships were warned via several newspapers, including *The Massachusetts Gazette*, "We are prepared, and shall not fail to pay them an unwelcome visit; by The Mohawks."

Step forward in time to 1860—the eve of the American Civil War—to another near monopoly supplier of historical significance: the U.S. cotton industry. At that time, the Southern states provided the majority of the cotton Britain imported. The South, wanting to secede from the Union, hoped to leverage Britain's high dependency on its cotton into formal diplomatic recognition of the Confederate States of America.

Figure 9.1. *Political Power from a Cotton Monopoly. In the mid-nineteenth century, the United States, specifically the Southern states, had a near monopoly in the cotton supplied to Great Britain. These states attempted to leverage this economic power into political power—trying to sway Great Britain to formally recognize the Confederate States of America. (Credit: modification of work by "ashleylovespizza"/Flickr Creative Commons)*

This leads us to the topic of this module: a firm that controls all (or nearly all) of the supply of a good or service—a monopoly. How do monopoly firms behave in the marketplace? Do they have "power?" Does this power potentially have unintended consequences? We'll return to this case at the end of the module to see how the tea and cotton monopolies influenced U.S. history.



Monopoly by Christopher Dombres, CC-BY.

Introduction to a Monopoly

There is a widespread belief that top executives at firms are the strongest supporters of market competition, but this belief is far from the truth. Think about it this way: If you very much wanted to win an Olympic gold medal, would you rather be far better than everyone else, or locked in competition with many athletes just as good as you are? Similarly, if you would like to attain a very high level of profits, would you rather manage a business with little or no competition, or struggle against many tough competitors who are trying to sell to your customers? By now, you might have read the module on Perfect Competition. In this module we explore the opposite extreme: monopoly.

If perfect competition is a market where firms have no market power and they simply respond to the market price, monopoly is a market with no competition at all, and firms have complete market power. In the case of *monopoly*, one firm produces all of the output in a market. Since a monopoly faces no significant competition, it can charge any price it wishes. While a monopoly, by definition, refers to a single firm, in practice the term is often used to describe a market in which one firm merely has a very high market share. This tends to be the definition that the U.S. Department of Justice uses.

Even though there are very few true monopolies in existence, we do deal with some of those few every day, often without realizing it: The U.S. Postal Service, your electric and garbage collection companies are a few examples. Some new drugs are produced by only one pharmaceutical firm—and no close substitutes for that drug may exist.

From the mid-1990s until 2004, the U.S. Department of Justice prosecuted the Microsoft Corporation for including Internet Explorer as the default web browser with its operating system. The Justice Department's argument was that, since Microsoft possessed an extremely high market share in the industry for operating systems, the inclusion of a free web browser constituted unfair competition to other browsers, such as Netscape Navigator. Since nearly everyone was using Windows, including Internet Explorer eliminated the incentive for consumers to explore other browsers and made it impossible for competitors to gain a foothold in the market. In 2013, the Windows system ran on more than 90% of the most commonly sold personal computers.

This module begins by describing how monopolies are protected from competition, including laws that prohibit competition, technological advantages, and certain configurations of demand and supply. It then discusses how a monopoly will choose its profit-maximizing quantity to produce and what price to charge. While a monopoly must be concerned about whether consumers will purchase its products or spend their money on something altogether different, the monopolist need not worry about the actions of other competing firms producing its products. As a result, a monopoly is not a price taker like a perfectly competitive firm, but instead exercises some power to choose its market price.

Self Check: Introduction to Monopolies

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/663

190. Outcome: Barriers to Entry

What you'll learn to do: define and explain the sources of barriers to entry

So how does someone go about creating a monopoly, anyway? In this section, you'll learn about the characteristics that lead to an environment where one business or firm could become a monopolist. Some monopolies form out of a government requirement, some because they control an essential resource, some because they are given an patent or have a copyright, or others because they intimidate competitors.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: How Monopolies Form: Barriers to Entry
- Self Check: Barriers to Entry

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

191. Reading: How Monopolies Form: Barriers to Entry

How Monopolies Form: Barriers to Entry

Because of the lack of competition, monopolies tend to earn significant economic profits. These profits should attract vigorous competition as described in Perfect Competition, and yet, because of one particular characteristic of monopoly, they do not. ***Barriers to entry*** are the legal, technological, or market forces that discourage or prevent potential competitors from entering a market. Barriers to entry can range from the simple and easily surmountable, such as the cost of renting retail space, to the extremely restrictive. For example, there are a finite number of radio frequencies available for broadcasting. Once the rights to all of them have been purchased, no new competitors can enter the market.

In some cases, barriers to entry may lead to monopoly. In other cases, they may limit competition to a few firms. Barriers may block entry even if the firm or firms currently in the market are earning profits. Thus, in markets with significant barriers to entry, it is *not* true that abnormally high profits will attract new firms, and that this entry of new firms will eventually cause the price to decline so that surviving firms earn only a normal level of profit in the long run.

There are two types of monopoly, based on the types of barriers to entry they exploit. One is ***natural monopoly***, where the barriers to entry are something other than legal prohibition. The other is ***legal monopoly***, where laws prohibit (or severely limit) competition.

NATURAL MONOPOLY

Economies of scale can combine with the size of the market to limit competition. Figure 9.2 presents a long-run average cost curve for the airplane manufacturing industry. It shows economies of scale up to an output of 8,000 planes per year and a price of P_0, then constant returns to scale from 8,000 to 20,000 planes per year, and diseconomies of scale at a quantity of production greater than 20,000 planes per year.

Now consider the market demand curve in the diagram, which intersects the long-run average cost (LRAC) curve at an output level of 6,000 planes per year and at a price P_1, which is higher than P_0. In this situation, the market has room for only one producer. If a second firm attempts to enter the market at a smaller size, say by producing a quantity of 4,000 planes, then its average costs will be higher than the existing firm, and it will be unable to compete. If the second firm attempts to enter the market at a larger size, like 8,000 planes per year, then it could produce at a lower average cost—but it could not sell all 8,000 planes that it produced because of insufficient demand in the market.

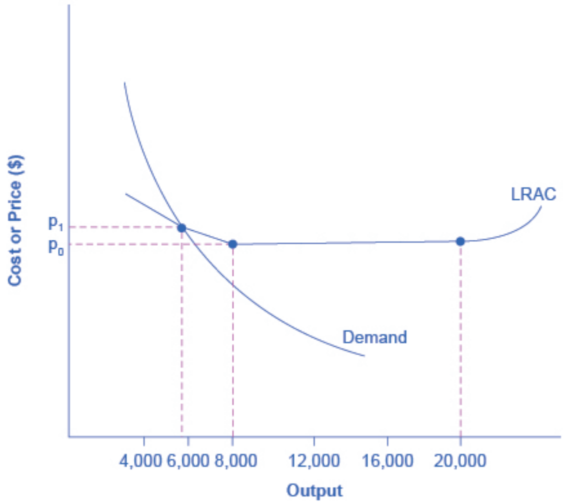


Figure 9.2. *Economies of Scale and Natural Monopoly. In this market, the demand curve intersects the long-run average cost (LRAC) curve at its downward-sloping part. A natural monopoly occurs when the quantity demanded is less than the minimum quantity it takes to be at the bottom of the long-run average cost curve.*

This situation, when economies of scale are large relative to the quantity demanded in the market, is called a natural monopoly. Natural monopolies often arise in industries where the marginal cost of adding an additional customer is very low, once the fixed costs of the overall system are in place. Once the main water pipes are laid through a neighborhood, the marginal cost of providing water service to another home is fairly low. Once electricity lines are installed through a neighborhood, the marginal cost of providing additional electrical service to one more home is very low. It would be costly and duplicative for a second water company to enter the market and invest in a whole second set of main water pipes, or for a second electricity company to enter the market and invest in a whole new set of electrical wires. These industries offer an example where, because of economies of scale, one producer can serve the entire market more efficiently than a number of smaller producers that would need to make duplicate physical capital investments.

A natural monopoly can also arise in smaller local markets for products that are difficult to transport. For example, cement production exhibits economies of scale, and the quantity of cement demanded in a local area may not be much larger than what a single plant can produce. Moreover, the costs of transporting cement over land are high, and so a cement plant in an area without access to water transportation may be a natural monopoly.

CONTROL OF A PHYSICAL RESOURCE

Another type of natural monopoly occurs when a company has control of a scarce physical resource. In the U.S. economy, one historical example of this pattern occurred when ALCOA—the Aluminum Company of America—controlled most of the supply of bauxite, a key mineral used in making aluminum. Back in the 1930s, when ALCOA controlled most of the bauxite, other firms were simply unable to produce enough aluminum to compete.

As another example, the majority of global diamond production is controlled by DeBeers, a multi-national company that has mining and production operations in South Africa, Botswana, Namibia, and Canada. It also has exploration activities on four continents, while directing a worldwide distribution network of rough diamonds. Though in recent years they have experienced growing competition, their impact on the rough diamond market is still considerable.

LEGAL MONOPOLY

For some products, the government erects barriers to entry by prohibiting or limiting competition. Under U.S. law, no organization but the U.S. Postal Service is legally allowed to deliver first-class mail. Many states or cities have laws or regulations that allow households a choice of only one electric company, one water company, and one company to pick up the garbage. Most legal monopolies are considered utilities—products necessary for everyday life—that are socially beneficial to have. As a consequence, the government allows producers to become regulated monopolies, to insure that an appropriate amount of these products is provided to consumers. Additionally, legal monopolies are often subject to economies of scale, so it makes sense to allow only one provider.

PROMOTING INNOVATION

Innovation takes time and resources to achieve. Suppose a company invests in research and development and finds the cure for the common cold. In this world of near ubiquitous information, other companies could take the formula, produce the drug, and because they did not incur the costs of research and development (R&D), undercut the price of the company that discovered the drug. Given this possibility, many firms would choose not to invest in research and development, and as a result, the world would have less innovation. To prevent this from happening, the Constitution of the United States specifies in Article I, Section 8: "The Congress shall have Power . . . To Promote the Progress of Science and Useful Arts, by securing for limited Times to Authors and Inventors the Exclusive Right to their Writings and Discoveries." Congress used this power to create the U.S. Patent and Trademark Office, as well as the U.S. Copyright Office. A *patent* gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time; in the United States, exclusive patent rights last for 20 years. The idea is to provide limited monopoly power so that innovative firms can recoup their investment in R&D, but then to allow other firms to produce the product more cheaply once the patent expires.

A ***trademark*** is an identifying symbol or name for a particular good, like Chiquita bananas, Chevrolet cars, or the Nike "swoosh" that appears on shoes and athletic gear. Roughly 1.9 million trademarks are registered with the U.S. government. A firm can renew a trademark over and over again, as long as it remains in active use.

A ***copyright***, according to the U.S. Copyright Office, "is a form of protection provided by the laws of the United States for 'original works of authorship' including literary, dramatic, musical, architectural, cartographic, choreographic, pantomimic, pictorial, graphic, sculptural, and audiovisual creations." No one can reproduce, display, or perform a copyrighted work without permission of the author. Copyright protection ordinarily lasts for the life of the author plus 70 years.

Roughly speaking, patent law covers inventions and copyright protects books, songs, and art. But in certain areas, like the invention of new software, it has been unclear whether patent or copyright protection should apply. There is also a body of law known as ***trade secrets***. Even if a company does not have a patent on an invention, competing firms are not allowed to steal their secrets. One famous trade secret is the formula for Coca-Cola, which is not protected under copyright or patent law, but is simply kept secret by the company.

Taken together, this combination of patents, trademarks, copyrights, and trade secret law is called ***intellectual property***, because it implies ownership over an idea, concept, or image, not a physical piece of property like a house or a car. Countries around the world have enacted laws to protect intellectual property, although the time periods and exact provisions of such laws vary across countries. There are ongoing negotiations, both through the World Intellectual Property Organization (WIPO) and through international treaties, to bring greater harmony to the intellectual property laws of different countries to determine the extent to which patents and copyrights in one country will be respected in other countries.

Government limitations on competition used to be even more common in the United States. For most of the twentieth century, only one phone company—AT&T—was legally allowed to provide local and long distance service. From the 1930s to the 1970s, one set of federal regulations limited which destinations airlines could choose to fly to and what fares they could charge; another set of regulations limited the interest rates that banks could pay to depositors; yet another specified what trucking firms could charge customers.

What products are considered utilities depends, in part, on the available technology. Fifty years ago, local and long distance telephone service was provided over wires. It did not make much sense to have multiple companies building multiple systems of wiring across towns and across the country. AT&T lost its monopoly on long distance service when the technology for providing phone service changed from wires to microwave and satellite transmission, so that multiple firms could use the same transmission mechanism. The same thing happened to local service, especially in recent years, with the growth in cellular phone systems.

The combination of improvements in production technologies and a general sense that the markets could provide services adequately led to a wave of ***deregulation***, starting in the late 1970s and continuing into the 1990s. This wave eliminated or reduced government restrictions on the firms that could enter, the prices that could be charged, and the quantities that could be produced in many industries, including telecommunications, airlines, trucking, banking, and electricity.

Around the world, from Europe to Latin America to Africa and Asia, many governments continue to control and limit competition in what those governments perceive to be key industries, including airlines, banks, steel companies, oil companies, and telephone companies.

LINK IT UP

Visit this website for examples of some pretty bizarre patents.

INTIMIDATING POTENTIAL COMPETITION

Businesses have developed a number of schemes for creating barriers to entry by deterring potential competitors from entering the market. One method is known as ***predatory pricing***, in which a firm uses the threat of sharp price cuts to discourage competition. Predatory pricing is a violation of U.S. antitrust law, but it is difficult to prove.

Consider a large airline that provides most of the flights between two particular cities. A new, small start-up airline decides to offer service between these two cities. The large airline immediately slashes prices on this route to the bone, so that the new entrant cannot make any money. After the new entrant has gone out of business, the incumbent firm can raise prices again.

After this pattern is repeated once or twice, potential new entrants may decide that it is not wise to try to compete. Small airlines often accuse larger airlines of predatory pricing: in the early 2000s, for example, ValuJet accused Delta of predatory pricing, Frontier accused United, and Reno Air accused Northwest. In late 2009, the American Booksellers Association, which represents independently owned and often smaller bookstores, accused Amazon, Wal-Mart, and Target of predatory pricing for selling new hardcover best-sellers at low prices.

In some cases, large advertising budgets can also act as a way of discouraging the competition. If the only way to launch a successful new national cola drink is to spend more than the promotional budgets of Coca-Cola and Pepsi Cola, not too many companies will try. A firmly established brand name can be difficult to dislodge.

SUMMING UP BARRIERS TO ENTRY

Table 9.1 lists the barriers to entry that have been discussed here. This list is not exhaustive, since firms have proved to be highly creative in inventing business practices that discourage competition. When barriers to entry exist, perfect competition is no longer a reasonable description of how an industry works. When barriers to entry are high enough, monopoly can result.

Table 9.1 Barriers to Entry

Barrier to Entry	Government Role?	Example
Natural monopoly	Government often responds with regulation (or ownership)	Water and electric companies
Control of a physical resource	No	DeBeers for diamonds
Legal monopoly	Yes	Post office, past regulation of airlines and trucking
Patent, trademark, and copyright	Yes, through protection of intellectual property	New drugs or software
Intimidating potential competitors	Somewhat	Predatory pricing; well-known brand names

Self Check: Barriers to Entry

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/664

192. Outcome: Revenue, Costs, Profit and Losses in Monopolies

What you'll learn to do: calculate and graph a monopoly's fixed, variable, average, marginal and total costs

We know that because a monopolist controls the market for a good or service, they get more say in how much they want to produce and what price to sell it at. In this outcome, you'll see how they make those decisions.

Here are some of the specific things you'll learn to do in this section:

- Measure variable and total costs as the area under the average variable and average total cost curves
- Calculate and graph the firm's average, marginal and total revenues
- Measure total revenues as the area under the average revenue curves
- Determine the profit maximizing output level and price; calculate and graphically illustrate where marginal revenue equals marginal costs
- Calculate and graphically illustrate profit and losses for a monopolist

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Choosing Output and Price
- Reading: Illustrating Monopoly Profits
- Self Check: Revenue, Costs, Profit and Losses in Monopolies

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

193. Reading: Choosing Output and Price

Demand Curves Perceived by a Perfectly Competitive Firm and by a Monopoly

Consider a monopoly firm, comfortably surrounded by barriers to entry so that it need not fear competition from other producers. How will this monopoly choose its profit-maximizing quantity of output, and what price will it charge? Profits for the monopolist, like any firm, will be equal to total revenues minus total costs. The pattern of costs for the monopoly can be analyzed within the same framework as the costs of a ***perfectly competitive firm***—that is, by using total cost, fixed cost, variable cost, marginal cost, average cost, and average variable cost. However, because a monopoly faces no competition, its situation and its decision process will differ from that of a perfectly competitive firm.

A perfectly competitive firm acts as a price taker, so its calculation of total revenue is made by taking the given market price and multiplying it by the quantity of output that the firm chooses. The demand curve *as it is perceived by a perfectly competitive firm* appears in Figure 9.3 (a). The flat perceived demand curve means that, from the viewpoint of the perfectly competitive firm, it could sell either a relatively low quantity like Ql or a relatively high quantity like Qh at the market price P.

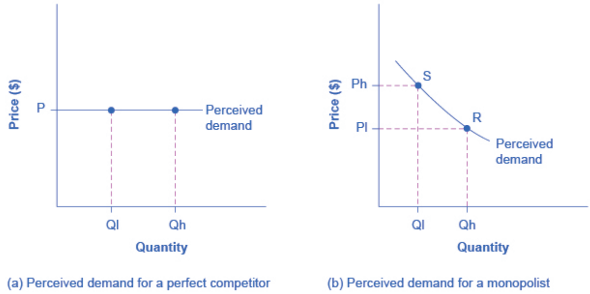


Figure 9.3. *The Perceived Demand Curve for a Perfect Competitor and a Monopolist. (a) A perfectly competitive firm perceives the demand curve that it faces to be flat. The flat shape means that the firm can sell either a low quantity (Ql) or a high quantity (Qh) at exactly the same price (P). (b) A monopolist perceives the demand curve that it faces to be the same as the market demand curve, which for most goods is downward-sloping. Thus, if the monopolist chooses a high level of output (Qh), it can charge only a relatively low price (Pl); conversely, if the monopolist chooses a low level of output (Ql), it can then charge a higher price (Ph). The challenge for the monopolist is to choose the combination of price and quantity that maximizes profits.*

What Defines the Market?

A monopoly is a firm that sells all or nearly all of the goods and services in a given market. But what defines the "market"?

In a famous 1947 case, the federal government accused the DuPont company of having a monopoly in the cellophane market, pointing out that DuPont produced 75% of the cellophane in the United States. DuPont countered that even though it had a 75% market share in cellophane, it had less than a 20% share of the "flexible packaging materials," which includes all other moistureproof papers, films, and foils. In 1956, after years of legal appeals, the U.S. Supreme Court held that the broader market definition was more appropriate, and the case against DuPont was dismissed.

Questions over how to define the market continue today. True, Microsoft in the 1990s had a dominant share of the software for computer operating systems, but in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 16% in 2000. The Greyhound bus company may have a near-monopoly on the market for intercity bus transportation, but it is only a small share of the market for intercity transportation if that market includes private cars, airplanes, and railroad service. DeBeers has a monopoly in diamonds, but it is a much smaller share of the total market for precious gemstones and an even smaller share of the total market for jewelry. A small town in the country may have only one gas station: is this gas station a "monopoly," or does it compete with gas stations that might be five, 10, or 50 miles away?

In general, if a firm produces a product without close substitutes, then the firm can be considered a monopoly producer in a single market. But if buyers have a range of similar—even if not identical—options available from other firms, then the firm is not a monopoly. Still, arguments over whether substitutes are close or not close can be controversial.

While a monopolist can charge *any* price for its product, that price is nonetheless constrained by demand for the firm's product. No monopolist, even one that is thoroughly protected by high barriers to entry, can require consumers to purchase its product. Because the monopolist is the only firm in the market, its demand curve is the same as the market demand curve, which is, unlike that for a perfectly competitive firm, downward-sloping.

Figure 9.3 illustrates this situation. The monopolist can either choose a point like R with a low price (Pl) and high quantity (Qh), or a point like S with a high price (Ph) and a low quantity (Ql), or some intermediate point. Setting the price too high will result in a low quantity sold, and will not bring in much revenue. Conversely, setting the price too low may result in a high quantity sold, but because of the low price, it will not bring in much revenue either. The challenge for the monopolist is to strike a profit-maximizing balance between the price it charges and the quantity that it sells. But why isn't the perfectly competitive firm's demand curve also the market demand curve?

WHAT IS THE DIFFERENCE BETWEEN PERCEIVED DEMAND AND MARKET DEMAND?

The demand curve as perceived by a perfectly competitive firm is not the overall market demand curve for that product. However, the firm's demand curve as perceived by a monopoly is the same as the market demand curve. The reason for the difference is that each perfectly competitive firm perceives the demand for its products in a market that includes many other firms; in effect, the demand curve perceived by a perfectly competitive firm is a tiny slice of the entire market demand curve. In contrast, a monopoly perceives demand for its product in a market where the monopoly is the only producer.

Total Cost and Total Revenue for a Monopolist

Profits for a monopolist can be illustrated with a graph of total revenues and total costs, as shown with the example of the hypothetical HealthPill firm in Figure 9.4. The total cost curve has its typical shape; that is, total costs rise and the curve grows steeper as output increases.

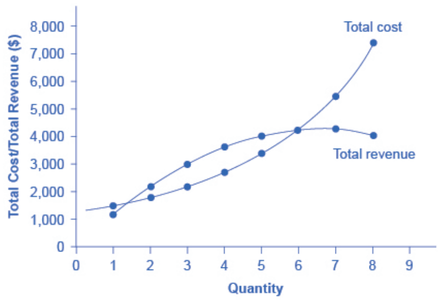


Figure 9.4. *Total Revenue and Total Cost for the HealthPill Monopoly Total revenue for the monopoly firm called HealthPill first rises, then falls. Low levels of output bring in relatively little total revenue, because the quantity is low. High levels of output bring in relatively less revenue, because the high quantity pushes down the market price. The total cost curve is upward-sloping. Profits will be highest at the quantity of output where total revenue is most above total cost. Of the choices in Table 9.2, the highest profits happen at an output of 4.*

The profit-maximizing level of output is not the same as the revenue-maximizing level of output, which should make sense, because profits take costs into account and revenues do not.

Table 9.2 Total Costs and Total Revenues of HealthPill

Quantity	*Total Cost*	*Quantity*	*Price*	*Total Revenue*	*Profit = Total Revenue − Total Cost*
1	1,500	1	1,200	1,200	−300
2	1,800	2	1,100	2,200	400
3	2,200	3	1,000	3,000	800
4	2,800	4	900	3,600	900
5	3,500	5	800	4,000	700
6	4,200	6	700	4,200	0
7	5,600	7	600	4,200	−1,400
8	7,400	8	500	4,000	−3,400

To calculate total revenue for a monopolist, start with the demand curve perceived by the monopolist. Table 9.2 shows quantities along the demand curve and the price at each quantity demanded, and then calculates total revenue by multiplying price times quantity at each level of output. (In this example, the output is given as 1, 2, 3, 4, and so on, for the sake of simplicity. If you prefer a dash of greater realism, you can imagine that these output levels and the corresponding prices are measured per 1,000 or 10,000 pills.) As the figure illustrates, total revenue for a monopolist rises, flattens out, and then falls. In this example, total revenue is highest at a quantity of 6 or 7.

Clearly, the total revenue for a monopolist is not a straight upward-sloping line, in the way that total revenue was for a perfectly competitive firm. The different total revenue pattern for a monopolist occurs because the quantity that a monopolist chooses to produce affects the market price, which was not true for a perfectly competitive firm. If the monopolist charges a very high price, then quantity demanded drops, and so total revenue is very low. If the monopolist charges a very low price, then, even if quantity demanded is very high, total revenue will not add up to much. At some intermediate level, total revenue will be highest.

However, the monopolist is not seeking to maximize revenue, but instead to earn the highest possible profit. Profits are calculated in the final row of the table. In the HealthPill example in Figure 9.4, the highest profit will occur at the quantity where total revenue is the farthest above total cost. Of the choices given in the table, the highest profits occur at an output of 4, where profit is 900.

Marginal Revenue and Marginal Cost for a Monopolist

In the real world, a monopolist often does not have enough information to analyze its entire total revenues or total costs curves; after all, the firm does not know exactly what would happen if it were to alter production dramatically. But a monopolist often has fairly reliable information about how changing output by small or moderate amounts will affect its marginal revenues and marginal costs, because it has had experience with such changes over time and because modest changes are easier to extrapolate from current experience. A monopolist can use information on ***marginal revenue*** and ***marginal cost*** to seek out the profit-maximizing combination of quantity and price.

The first four columns of Table 9.3 use the numbers on total cost from the HealthPill example in the previous exhibit and calculate marginal cost and average cost. This monopoly faces a typical U-shaped average cost curve and upward-sloping marginal cost curve, as shown in Figure 9.5. The second four columns of Table 9.3 use the total revenue information from the previous exhibit and calculate marginal revenue.

Notice that marginal revenue is zero at a quantity of 7, and turns negative at quantities higher than 7. It may seem counterintuitive that marginal revenue could ever be zero or negative: after all, does an increase in quantity sold not always mean more revenue? For a perfect competitor, each additional unit sold brought a positive marginal revenue, because marginal revenue was equal to the given market price. But a monopolist can sell a larger quantity and see a decline in *total revenue*. When a monopolist increases sales by one unit, it gains some marginal revenue from selling that extra unit, but also loses some marginal revenue because every other unit must now be sold at a lower price. As the quantity sold becomes higher, the drop in price affects a greater quantity of sales, eventually causing a situation where more sales cause marginal revenue to be negative.

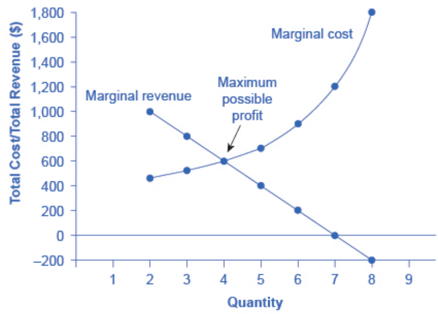


Figure 9.5. *Marginal Revenue and Marginal Cost for the HealthPill Monopoly. For a monopoly like HealthPill, marginal revenue decreases as additional units are sold. The marginal cost curve is upward-sloping. The profit-maximizing choice for the monopoly will be to produce at the quantity where marginal revenue is equal to marginal cost: that is, MR = MC. If the monopoly produces a lower quantity, then MR > MC at those levels of output, and the firm can make higher profits by expanding output. If the firm*

produces at a greater quantity, then MC > MR, and the firm can make higher profits by reducing its quantity of output.

Table 9.3 Costs and Revenues of HealthPill

Cost Information				**Revenue Information**			
Quantity	**Total Cost**	**Marginal Cost**	**Average Cost**	**Quantity**	**Price**	**Total Revenue**	**Total Revenue**
1	1,500	1,500	1,500	1	1,200	1,200	1,200
2	1,800	300	900	2	1,100	2,200	1,000
3	2,200	400	733	3	1,000	3,000	800
4	2,800	600	700	4	900	3,600	600
5	3,500	700	700	5	800	4,000	400
6	4,400	900	733	6	700	4,200	200
7	5,600	1,200	800	7	600	4,200	0
8	7,400	1,800	925	8	500	4,000	−200

A monopolist can determine its profit-maximizing price and quantity by analyzing the marginal revenue and marginal costs of producing an extra unit. If the marginal revenue exceeds the marginal cost, then the firm should produce the extra unit.

For example, at an output of 3 in Figure 9.5, marginal revenue is 800 and marginal cost is 400, so producing this unit will clearly add to overall profits. At an output of 4, marginal revenue is 600 and marginal cost is 600, so producing this unit still means overall profits are unchanged. However, expanding output from 4 to 5 would involve a marginal revenue of 400 and a marginal cost of 700, so that fifth unit would actually reduce profits. Thus, the monopoly can tell from the marginal revenue and marginal cost that of the choices given in the table, the profit-maximizing level of output is 4.

Indeed, the monopoly could seek out the profit-maximizing level of output by increasing quantity by a small amount, calculating marginal revenue and marginal cost, and then either increasing output as long as marginal revenue exceeds marginal cost or reducing output if marginal cost exceeds marginal revenue. This process works without any need to calculate total revenue and total cost. Thus, a profit-maximizing monopoly should follow the rule of producing up to the quantity where marginal revenue is equal to marginal cost—that is, MR = MC.

If you find it counterintuitive that producing where marginal revenue equals marginal cost will maximize profits, working through the numbers will help.

Step 1. Remember that marginal cost is defined as the change in total cost from producing a small amount of additional output.

$$MC = \frac{\text{change in total cost}}{\text{change in quantity produced}}$$

Step 2. Note that in Table 9.3 , as output increases from 1 to 2 units, total cost increases from \$1500 to \$1800. As a result, the marginal cost of the second unit will be:

$$\begin{aligned} MC &= \frac{\$1800 - \$1500}{1} \\ &= \$300 \end{aligned}$$

Step 3. Remember that, similarly, marginal revenue is the change in total revenue from selling a small amount of additional output.

$$MR = \frac{\text{change in total revenue}}{\text{change in quantity sold}}$$

Step 4. Note that in Table 9.3 , as output increases from 1 to 2 units, total revenue increases from \$1200 to \$2200. As a result, the marginal revenue of the second unit will be:

$$\begin{aligned} MR &= \frac{\$2200 - \$1200}{1} \\ &= \$1000 \end{aligned}$$

Table 9.4 repeats the marginal cost and marginal revenue data from Table 9.3, and adds two more columns: *Marginal profit* is the profitability of each additional unit sold. It is defined as marginal revenue minus marginal cost. Finally, total profit is the sum of marginal profits. As long as marginal profit is positive, producing more output will increase total profits. When marginal profit turns negative, producing more output will decrease total profits. Total profit is maximized where marginal revenue equals marginal cost. In this example, maximum profit occurs at 4 units of output.

A perfectly competitive firm will also find its profit-maximizing level of output where MR = MC. The key difference with a perfectly competitive firm is that in the case of perfect competition, marginal revenue is equal to price (MR = P), while for a monopolist, marginal revenue is not equal to the price, because changes in quantity of output affect the price.

Table 9.4 Marginal Revenue, Marginal Cost, Marginal and Total Profit

Quantity	Marginal Revenue	Marginal Cost	Marginal Profit	Total Profit
1	1,200	1,500	–300	–300
2	1,000	300	700	400
3	800	400	400	800
4	600	600	0	800
5	400	700	–300	500
6	200	900	–700	–200
7	0	1,200	–1,200	–1,400

194. Reading: Illustrating Monopoly Profits

Illustrating Monopoly Profits

It is straightforward to calculate profits of given numbers for total revenue and total cost. However, the size of monopoly profits can also be illustrated graphically with Figure 9.6, which takes the marginal cost and marginal revenue curves from the previous exhibit and adds an average cost curve and the monopolist's perceived demand curve.

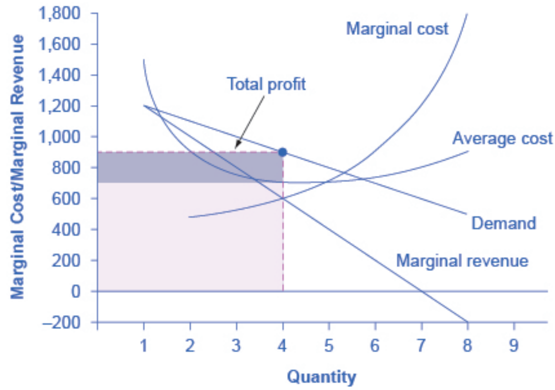


Figure 9.6. *Illustrating Profits at the HealthPill Monopoly. This figure begins with the same marginal revenue and marginal cost curves from the HealthPill monopoly presented in Figure 9.5. It then adds an average cost curve and the demand curve faced by the monopolist. The HealthPill firm first chooses the quantity where MR = MC; in this example, the quantity is 4. The monopolist then decides what price to charge by looking at the demand curve it faces. The large box, with quantity on the horizontal axis and marginal revenue on the vertical axis, shows total revenue for the firm. Total costs for the firm are shown by the lighter-shaded box, which is quantity on the horizontal axis and marginal cost of production on the vertical axis. The large total revenue box minus the smaller total cost box leaves the darkly shaded box that shows total profits. Since the price charged is above average cost, the firm is earning positive profits.*

The three-step process where a monopolist selects the profit-maximizing quantity to produce, decides what price to charge, and then determines total revenue, total cost and profit. These steps include:

Step 1: The Monopolist Determines Its Profit-Maximizing Level of Output

The firm can use the points on the demand curve D to calculate total revenue, and then, based on total revenue, calculate its marginal revenue curve. The profit-maximizing quantity will occur where MR = MC—or at the last possible point before marginal costs start exceeding marginal revenue. On Figure 9.6, MR = MC occurs at an output of 4.

Step 2: The Monopolist Decides What Price to Charge

The monopolist will charge what the market is willing to pay. A dotted line drawn straight up from the profit-maximizing quantity to the demand curve shows the profit-maximizing price. This price is above the average cost curve, which shows that the firm is earning profits.

Step 3: Calculate Total Revenue, Total Cost, and Profit

Total revenue is the overall shaded box, where the width of the box is the quantity being sold and the height is the price. In Figure 9.6, the bottom part of the shaded box, which is shaded more lightly, shows total costs; that is, quantity on the horizontal axis multiplied by average cost on the vertical axis. The larger box of total revenues minus the smaller box of total costs will equal profits, which is shown by the darkly shaded box.

This three-step process is illustrated in Figure 9.7.

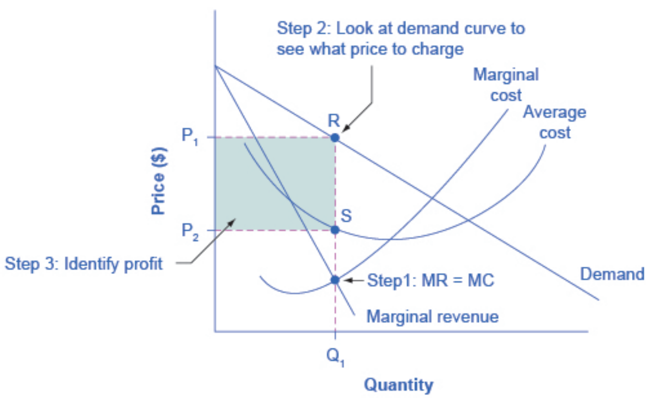


Figure 9.7. *How a Profit-Maximizing Monopoly Decides Price. In Step 1, the monopoly chooses the profit-maximizing level of output Q1, by choosing the quantity where MR = MC. In Step 2, the monopoly decides how much to charge for output level Q1 by drawing a line straight up from Q1 to point R on its perceived demand curve. Thus,the monopoly will charge a price (P1). In Step 3, the monopoly identifies its profit. Total revenue will be Q1 multiplied by P1. Total cost will be Q1 multiplied by the average cost of producing Q1, which is shown by point S on the average cost curve to be P2. Profits will be the total revenue rectangle minus the total cost rectangle, shown by the shaded zone in the figure.*

In a perfectly competitive market, the forces of entry would erode this profit in the long run. But a monopolist is protected by barriers to entry. In fact, one telltale sign of a possible monopoly is when a firm earns profits year after year, while doing more or less the same thing, without ever seeing those profits eroded by increased competition.

Why is a monopolist's marginal revenue always less than the price?

The marginal revenue curve for a monopolist always lies beneath the market demand curve. To understand why, think about increasing the quantity along the demand curve by one unit, so that you take one step down the demand curve to a slightly higher quantity but a slightly lower price. A demand curve is not sequential: It is not that first we sell Q1 at a higher price, and then we sell Q2 at a lower price. Rather, a demand curve is conditional: If we charge the higher price, we would sell Q1. If, instead, we charge a lower price (on all the units that we sell), we would sell Q2.

So when we think about increasing the quantity sold by one unit, marginal revenue is affected in two ways. First, we sell one additional unit at the new market price. Second, all the previous units, which could have been sold at the higher price, now sell for less. Because of the lower price on all units sold, the marginal revenue of selling a unit is less than the price of that unit—and the marginal revenue curve is below the demand curve.

Tip: For a straight-line demand curve, the marginal revenue curve equals price at the lowest level of output. (Graphically, MR and demand have the same vertical axis.) As output increases, marginal revenue decreases twice as fast as demand, so that the horizontal intercept of MR is halfway to the horizontal intercept of demand. You can see this in the Figure 9.8.

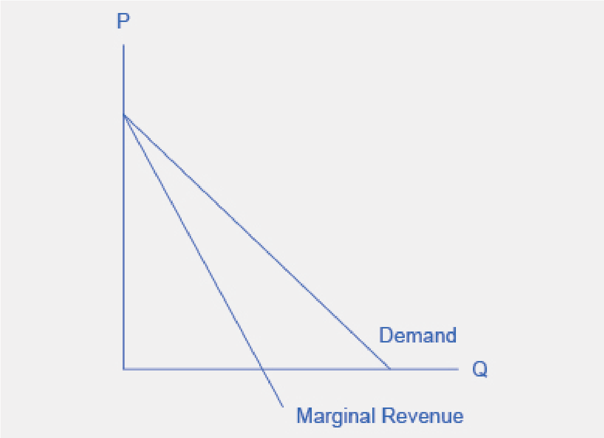


Figure 9.8. *The Monopolist's Marginal Revenue Curve versus Demand Curve Because the market demand curve is conditional, the marginal revenue curve for a monopolist lies beneath the demand curve.*

Watch the video below to see these monopoly profits illustrated on a graph.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=222#oembed-1

Self Check: Revenue, Costs, Profit and Losses in Monopolies

Answer the question(s) below to see how well you understand the

topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/769

195. Outcome: Inefficiencies in Monopolies

What you'll learn to do: explain why a monopoly is inefficient using deadweight loss

In this outcome, you'll see why monopolies are inefficient. They produce less, at a higher average cost, and charge a higher price than would a combination of firms in a perfectly competitive industry. Because of this, they are neither allocatively or productively efficient.

Here are some of the specific things you'll learn to do in this section:

- Differentiate between a single price monopolist and a price discriminating monopolist

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: The Inefficiency of Monopoly
- Reading: Monopolies and Deadweight Loss
- Reading: Price Discrimination
- Self Check: Inefficiencies in Monopolies

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

196. Reading: The Inefficiency of Monopoly

The Inefficiency of Monopoly

Most people criticize monopolies because they charge too high a price, but what economists object to is that monopolies do not supply enough output to be allocatively efficient. To understand why a monopoly is inefficient, it is useful to compare it with the benchmark model of perfect competition.

Allocative efficiency is a social concept. It refers to producing the optimal quantity of some output, the quantity where the marginal benefit to society of one more unit just equals the marginal cost. The rule of profit maximization in a world of perfect competition was for each firm to produce the quantity of output where P = MC, where the price (P) is a measure of how much buyers value the good and the marginal cost (MC) is a measure of what marginal units cost society to produce. Following this rule assures allocative efficiency. If P > MC, then the marginal benefit to society (as measured by P) is greater than the marginal cost to society of producing additional units, and a greater quantity should be produced. But in the case of monopoly, price is always greater than marginal cost at the profit-maximizing level of output, as can be seen by looking back at Figure 9.6. Thus, consumers will suffer from a monopoly because a lower quantity will be sold in the market, at a higher price, than would have been the case in a perfectly competitive market.

The problem of inefficiency for monopolies often runs even deeper than these issues, and also involves incentives for efficiency over longer periods of time. There are counterbalancing incentives here. On one side, firms may strive for new inventions and new intellectual property because they want to become monopolies and earn high profits—at least for a few years until the competition catches up. In this way, monopolies may come to exist because of competitive pressures on firms. However, once a barrier to entry is in place, a monopoly that does not need to fear competition can just produce the same old products in the same old way—while still ringing up a healthy rate of profit. John Hicks, who won the Nobel Prize for economics in 1972, wrote in 1935: “The best of all monopoly profits is a quiet life.” He did not mean the comment in a complimentary way. He meant that monopolies may bank their profits and slack off on trying to please their customers.

When AT&T provided all of the local and long-distance phone service in the United States, along with manufacturing most of the phone equipment, the payment plans and types of phones did not change much. The old joke was that you could have any color phone you wanted, as long as it was black. But in 1982, AT&T was split up by government litigation into a number of local phone companies, a long-distance phone company, and a phone equipment manufacturer. An explosion of innovation followed. Services like call waiting, caller ID, three-way calling, voice mail though the phone company, mobile phones, and wireless connections to the Internet all became available. A wide range of payment plans was offered, as well. It was no longer true that all phones were black; instead, phones came in a wide variety of shapes and colors. The end of the telephone monopoly brought lower prices, a greater quantity of services, and also a wave of innovation aimed at attracting and pleasing customers.

THE REST IS HISTORY

Earlier in the module, the East India Company and the Confederate States were presented as a monopoly or near monopoly provider of a good. Nearly every American schoolchild knows the result of the 'unwelcome visit' the 'Mohawks' bestowed upon Boston Harbor's tea-bearing ships—the Boston Tea Party. Regarding the cotton industry, we also know Great Britain remained neutral during the Civil War, taking neither side during the conflict.

Did the monopoly nature of these businesses have unintended and historical consequences? Might the American Revolution have been deterred, if the East India Company had sailed the tea-bearing ships back to England? Might the southern states have made different decisions had they not been so confident "King Cotton" would force diplomatic recognition of the Confederate States of America? Of course, it is not possible to definitively answer these questions; after all we cannot roll back the clock and try a different scenario. We can, however, consider the monopoly nature of these businesses and the roles they played and hypothesize about what might have occurred under different circumstances.

Perhaps if there had been legal free tea trade, the colonists would have seen things differently; there was smuggled Dutch tea in the colonial market. If the colonists had been able to freely purchase Dutch tea, they would have paid lower prices and avoided the tax.

What about the cotton monopoly? With one in five jobs in Great Britain depending on Southern cotton and the Confederate States nearly the sole provider of that cotton, why did Great Britain remain neutral during the Civil War? At the beginning of the war, Britain simply drew down massive stores of cotton. These stockpiles lasted until near the end of 1862. Why did Britain not recognize the Confederacy at that point? Two reasons: The Emancipation Proclamation and new sources of cotton. Having outlawed slavery throughout the United Kingdom in 1833, it was politically impossible for Great Britain, empty cotton warehouses or not, to recognize, diplomatically, the Confederate States. In addition, during the two years it took to draw down the stockpiles, Britain expanded cotton imports from India, Egypt, and Brazil.

Monopoly sellers often see no threats to their superior marketplace position. In these examples did the power of the monopoly blind the decision makers to other possibilities? Perhaps. But, as they say, the rest is history.

197. Reading: Monopolies and Deadweight Loss

Monopoly and Efficiency

The fact that price in monopoly exceeds marginal cost suggests that the monopoly solution violates the basic condition for economic efficiency, that the price system must confront decision makers with all of the costs and all of the benefits of their choices. Efficiency requires that consumers confront prices that equal marginal costs. Because a monopoly firm charges a price greater than marginal cost, consumers will consume less of the monopoly's good or service than is economically efficient.

To contrast the efficiency of the perfectly competitive outcome with the inefficiency of the monopoly outcome, imagine a perfectly competitive industry whose solution is depicted in Figure 10.7 "Perfect Competition, Monopoly, and Efficiency". The short-run industry supply curve is the summation of individual marginal cost curves; it may be regarded as the marginal cost curve for the industry. A perfectly competitive industry achieves equilibrium at point C, at price P_C and quantity Q_C.

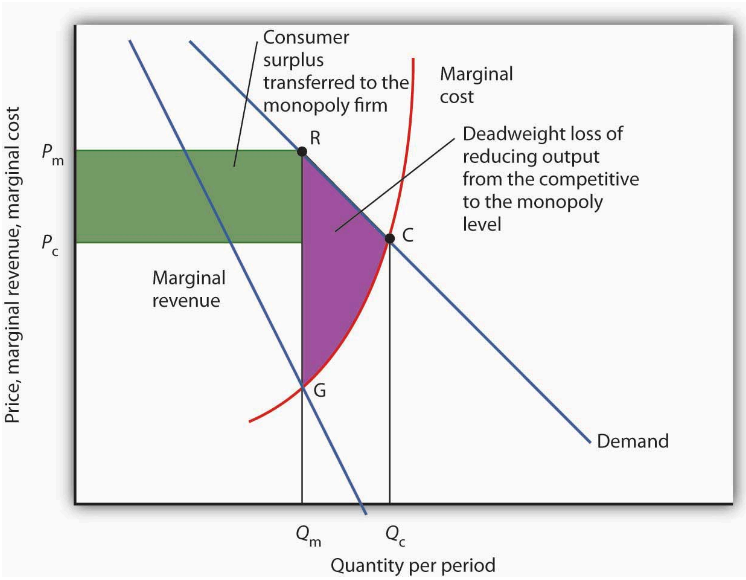


Figure 10.7 Perfect Competition, Monopoly, and Efficiency.

Given market demand and marginal revenue, we can compare the behavior of a monopoly to that of a perfectly competitive industry. The marginal cost curve may be thought of as the supply curve of a perfectly competitive industry. The perfectly competitive industry produces quantity Q_c and sells the output at price P_c. The monopolist restricts output to Q_m and raises the price to P_m.

Reorganizing a perfectly competitive industry as a monopoly results in a deadweight loss to society given by the shaded area GRC. It also transfers a portion of the consumer surplus earned in the competitive case to the monopoly firm.

Now, suppose that all the firms in the industry merge and a government restriction prohibits entry by any new firms. Our perfectly competitive industry is now a monopoly. Assume the monopoly continues to have the same marginal cost and demand curves that the competitive industry did. The monopoly firm faces the same market demand curve, from which it derives its marginal revenue curve. It maximizes profit at output Q_m and charges price P_m. Output is lower and price higher than in the competitive solution.

Society would gain by moving from the monopoly solution at Q_m to the competitive solution at Q_c. The benefit to consumers would be given by the area under the demand curve between Q_m and Q_c; it is the area Q_mRCQ_c. An increase in output, of course, has a cost. Because the marginal cost curve measures the cost of each additional unit, we can think of the area under the marginal cost curve over some range of output as measuring the total cost of that output. Thus, the total cost of increasing output from Q_m to Q_c is the area under the marginal cost curve over that range—the area Q_mGCQ_c. Subtracting this cost from the benefit gives us the net gain of moving from the monopoly to the competitive solution; it is the shaded area GRC. That is the potential gain from moving to the efficient solution. The area GRC is a deadweight loss.

198. Reading: Price Discrimination

Price Discrimination

Throughout this text up to this point, we have assumed that firms sold all units of output at the same price. In some cases, however, firms can charge different prices to different consumers. If such an opportunity exists, the firm can increase profits further.

When a firm charges different prices for the same good or service to different consumers, even though there is no difference in the cost to the firm of supplying these consumers, the firm is engaging in price discrimination. Except for a few situations of price discrimination that have been declared illegal, such as manufacturers selling their goods to distributors at different prices when there are no differences in cost, price discrimination is generally legal.

The potential for price discrimination exists in all market structures except perfect competition. As long as a firm faces a downward-sloping demand curve and thus has some degree of monopoly power, it may be able to engage in price discrimination. But monopoly power alone is not enough to allow a firm to price discriminate. Monopoly power is one of three conditions that must be met:

1. **A Price-Setting Firm** The firm must have some degree of monopoly power—it must be a price setter. A price-taking firm can only take the market price as given—it is not in a position to make price choices of any kind. Thus, firms in perfectly competitive markets will not engage in price discrimination. Firms in monopoly, monopolistically competitive, or

oligopolistic markets may engage in price discrimination.

2. **Distinguishable Customers** The market must be capable of being fairly easily segmented—separated so that customers with different elasticities of demand can be identified and treated differently.
3. **Prevention of Resale** The various market segments must be isolated in some way from one another to prevent customers who are offered a lower price from selling to customers who are charged a higher price. If consumers can easily resell a product, then discrimination is unlikely to be successful. Resale may be particularly difficult for certain services, such as dental checkups.

Examples of Price Discrimination

Examples of price discrimination abound. Senior citizens and students are often offered discount fares on city buses. Children receive discount prices for movie theater tickets and entrance fees at zoos and theme parks. Faculty and staff at colleges and universities might receive discounts at the campus bookstore. Airlines give discount prices to customers who are willing to stay over a Saturday night. Physicians might charge wealthy patients more than poor ones. People who save coupons are able to get discounts on many items. In all these cases a firm charges different prices to different customers for what is essentially the same product.

Not every instance of firms charging different prices to different customers constitutes price discrimination. Differences in prices may reflect different costs associated with providing the product. One buyer might require special billing practices, another might require delivery on a particular day of the week, and yet another might require special packaging. Price differentials based on differences in production costs are not examples of price discrimination.

Why would a firm charge different prices to different consumers? The answer can be found in the marginal decision rule and in the relationship between marginal revenue and elasticity.

Suppose an airline has found that its long-run profit-maximizing solution for a round-trip flight between Minneapolis and Cleveland, when it charges the same price to all passengers, is to carry 300 passengers at $200 per ticket. The airline has a degree of monopoly power, so it faces a downward-sloping demand curve. The airline has noticed that there are essentially two groups of customers on each flight: people who are traveling for business reasons and people who are traveling for personal reasons (visiting family or friends or taking a vacation). We will call this latter group "tourists." Of the 300 passengers, 200 are business travelers and 100 are tourists. The airline's revenue from business travelers is therefore currently $40,000 ($200 times 200 business travelers) and from tourists is currently $20,000 ($200 times 100 tourists).

It seems likely that the price elasticities of demand of these two groups for a particular flight will differ. Tourists may have a wide range of substitutes: They could take their trips at a different time, they could vacation in a different area, or they could easily choose not to go at all. Business travelers, however, might be attending meetings or conferences at a particular time and in a particular city. They have options, of course, but the range of options is likely to be more limited than the range of options facing tourists. Given all this, tourists are likely to have relatively more price elastic demand than business travelers for a particular flight.

The difference in price elasticities suggests the airline could increase its profit by adjusting its pricing. To simplify, suppose that at a price of about $200 per ticket, demand by tourists is relatively price elastic and by business travelers is relatively less price elastic. It is plausible that the marginal cost of additional passengers is likely to be quite low, since the number of crew members will not vary and no food is served on short flights. Thus, if the airline can increase its revenue, its profits will increase. Suppose the airline lowers the price for tourists to $190. Suppose that the lower price encourages 10 more tourists to take the flight. Of course, the airline cannot charge different prices to different tourists; rather it charges $190 to all, now 110, tourists. Still, the airline's revenue from tourist passengers increases from $20,000 to $20,900 ($190 times 110 tourists). Suppose it charges $250 to its business travelers. As a result, only 195 business travelers take the flight. The airline's revenue from business travelers still rises from $40,000 to $48,750 ($250 times 195 business travelers). The airline will continue to change the mix of passengers, and increase the number of passengers, so long as doing so increases its profit. Because tourist demand is relatively price elastic, relatively small reductions in price will attract relatively large numbers of additional tourists. Because business demand is relatively less elastic, relatively large increases in price will discourage relatively small numbers of business travelers from making the trip. The airline will continue to reduce its price to tourists and raise its price to business travelers as long as it gains profit from doing so.

Of course, the airline can impose a discriminatory fare structure only if it can distinguish tourists from business travelers. Airlines typically do this by looking at the travel plans of their customers. Trips that involve a stay over a weekend, for example, are more likely to be tourist related, whereas trips that begin and end during the workweek are likely to be business trips. Thus, airlines charge much lower fares for trips that extend through a weekend than for trips that begin and end on weekdays.

In general, price-discrimination strategies are based on differences in price elasticity of demand among groups of customers and the differences in marginal revenue that result. A firm will seek a price structure that offers customers with more elastic demand a lower price and offers customers with relatively less elastic demand a higher price.

It is always in the interest of a firm to discriminate. Yet most of the goods and services that we buy are not offered on a discriminatory basis. A grocery store does not charge a higher price for vegetables to vegetarians, whose demand is likely to be less elastic than that of its omnivorous customers. An audio store does not charge a different price for Pearl Jam's compact disks to collectors seeking a complete collection than it charges to casual fans who could easily substitute a disk from another performer. In these cases, firms lack a mechanism for knowing the different demands of their customers and for preventing resale.

Monopolies are usually considered to be inefficient, but if allowed to price discriminate they can (in a perfect world) be very efficient. Of course, unregulated monopolies are illegal in the U.S. and regulated monopolies generally are not allowed to price discriminate. For this reason, price discrimination is really a practice of imperfectly competitive firms (oligopolies and monopolistically competitive firms).

Self Check: Inefficiencies in Monopolies

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the three Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/770

199. Outcome: Controlling Monopolies

What you'll learn to do: analyze different strategies to control monopolies, including natural monopolies

In this section, we will learn about what measures the government takes to restrict and regulate monopolies. You'll learn more about corporate mergers, regulating anti-competitive behavior, natural monopolies, and the Great Deregulation Experiment.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Monopoly and Antitrust Policy
- Reading: Corporate Mergers
- Reading: Regulating Anticompetitive Behavior
- Reading: Regulating Natural Monopolies
- Reading: The Great Deregulation Experiment
- Self Check: Controlling Monopolies

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

200. Reading: Monopoly and Antitrust Policy

MORE THAN COOKING, HEATING, AND COOLING

If you live in the United States, there is a slightly better than 50–50 chance your home is heated and cooled using natural gas. You may even use natural gas for cooking. However, those uses are not the primary uses of natural gas in the U.S. In 2012, according to the U.S. Energy Information Administration, home heating, cooling, and cooking accounted for just 18% of natural gas usage. What accounts for the rest? The greatest uses for natural gas are the generation of electric power (39%) and in industry (30%). Together these three uses for natural gas touch many areas of our lives, so why would there be any opposition to a merger of two natural gas firms? After all, a merger could mean increased efficiencies and reduced costs to people like you and me.

In October 2011, Kinder Morgan and El Paso Corporation, two natural gas firms, announced they were merging. The announcement stated the combined firm would link "nearly every major production region with markets," cut costs by "eliminating duplication in pipelines and other assets," and that "the savings could be passed on to consumers."

Figure 11.1. *Oligopoly versus Competitors in the Marketplace. Large corporations, such as the natural gas producer Kinder Morgan, can bring economies of scale to the marketplace. Will that benefit consumers? Or is more competition better for consumers? (Credit: modification of work by Derrick Coetzee/ Flickr Creative Commons)*

The objection? The $21.1 billion deal would give Kinder Morgan control of more than 80,000 miles of pipeline, making the new firm the third largest energy producer in North America. As the third largest energy producer, policymakers and the public wondered whether the cost savings really would be passed on to consumers, or would the merger give Kinder Morgan a strong oligopoly position in the natural gas marketplace?

That brings us to the central question this module poses: What should the balance be between corporate size and a larger number of competitors in a marketplace? We will also consider what role the government should play in this balancing act.

Introduction to Monopoly and Antitrust Policy

The previous modules on the theory of the firm identified three important lessons: First, that competition, by providing consumers with lower prices and a variety of innovative products, is a good thing; second, that large-scale production can dramatically lower average costs; and third, that markets in the real world are rarely perfectly competitive. As a consequence, government policymakers must determine how much to intervene to balance the potential benefits of large-scale production against the potential loss of competition that can occur when businesses grow in size, especially through mergers.

For example, in 2006, AT&T and BellSouth, two telecommunications companies, wished to merge into a single firm. In the year before the merger, AT&T was the 121st largest company in the country when ranked by sales, with $44 billion in revenues and 190,000 employees. BellSouth was the 314th largest company in the country, with $21 billion in revenues and 63,000 employees.

The two companies argued that the merger would benefit consumers, who would be able to purchase better telecommunications services at a cheaper price because the newly created firm would be able to produce more efficiently by taking advantage of economies of scale and eliminating duplicate investments. However, a number of activist groups like the Consumer Federation of America and Public Knowledge expressed fears that the merger would reduce competition and lead to higher prices for consumers for decades to come. In December 2006, the federal government allowed the merger to proceed. By 2009, the new post-merger AT&T was the eighth largest company by revenues in the United States, and by that measure the largest telecommunications company in the world. Economists have spent – and will still spend – years trying to determine whether the merger of AT&T and BellSouth, as well as other smaller mergers of telecommunications companies at about this same time, helped consumers, hurt them, or did not make much difference.

This module discusses public policy issues about competition. How can economists and governments determine when mergers of large companies like AT&T and BellSouth should be allowed and when they should be blocked? The government also plays a role in policing anticompetitive behavior other than mergers, like prohibiting certain kinds of contracts that might restrict competition. In the case of natural monopoly, however, trying to preserve competition probably will not work very well, and so government will often resort to regulation of price and/or quantity of output. In recent decades, there has been a global trend toward less government intervention in the price and output decisions of businesses.

201. Reading: Corporate Mergers

Regulations for Approving Mergers

A corporate ***merger*** occurs when two formerly separate firms combine to become a single firm. When one firm purchases another, it is called an ***acquisition***. An acquisition may not look just like a merger, since the newly purchased firm may continue to be operated under its former company name. Mergers can also be lateral, where two firms of similar sizes combine to become one. However, both mergers and acquisitions lead to two formerly separate firms being under common ownership, and so they are commonly grouped together.

Since a merger combines two firms into one, it can reduce the extent of competition between firms. Therefore, when two U.S. firms announce a merger or acquisition where at least one of the firms is above a minimum size of sales (a threshold that moves up gradually over time, and was at $70.9 million in 2013), or certain other conditions are met, they are required under law to notify the U.S. Federal Trade Commission (FTC). The left-hand panel of Figure 11.2 (a) shows the number of mergers submitted for review to the FTC each year from 1999 to 2012. Mergers were very high in the late 1990s, diminished in the early 2000s, and then rebounded somewhat in a cyclical fashion. The right-hand panel of Figure 11.2 (b) shows the distribution of those mergers submitted for review in 2012 as measured by the size of the transaction. It is important to remember that this total leaves out many small mergers under $50 million, which only need to be reported in certain limited circumstances. About a quarter of all reported merger and acquisition transactions in 2012 exceeded $500 million, while about 11 percent exceeded $1 billion.

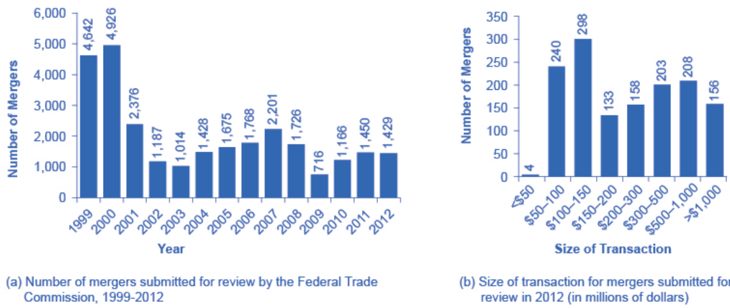


Figure 11.2. *Number and Size of Mergers. (a) The number of mergers in 1999 and 2000 were relatively high compared to the annual numbers seen from 2001–2012. While 2001 and 2007 saw a high number of mergers, these were still only about half the number of mergers in 1999 and 2000. (b) In 2012, the greatest number of mergers submitted for review was for transactions between $100 and $150 million.*

The laws that give government the power to block certain mergers, and even in some cases to break up large firms into smaller ones, are called ***antitrust laws***. Before a large merger happens, the antitrust regulators at the FTC and the U.S. Department of Justice can allow the merger, prohibit it, or allow it if certain conditions are met. One common condition is that the merger will be allowed if the firm agrees to sell off certain parts. For example, in 2006, Johnson & Johnson bought the Pfizer's "consumer health" division, which included well-known brands like Listerine mouthwash and Sudafed cold medicine. As a condition of allowing the merger, Johnson & Johnson was required to sell off six brands to other firms, including Zantac® heartburn relief medication, Cortizone anti-itch cream, and Balmex diaper rash medication, to preserve a greater degree of competition in these markets.

The U.S. government approves most proposed mergers. In a market-oriented economy, firms have the freedom to make their own choices. Private firms generally have the freedom to:

- expand or reduce production
- set the price they choose
- open new factories or sales facilities or close them
- hire workers or to lay them off
- start selling new products or stop selling existing ones

If the owners want to acquire a firm or be acquired, or to merge with another firm, this decision is just one of many that firms are free to make. In these conditions, the managers of private firms will sometimes make mistakes. They may close down a factory which, it later turns out, would have been profitable. They may start selling a product that ends up losing money. A merger between two companies can sometimes lead to a clash of corporate personalities that makes both firms worse off. But the fundamental belief behind a market-oriented economy is that firms, not governments, are in the best position to know if their actions will lead to attracting more customers or producing more efficiently.

Indeed, government regulators agree that most mergers are beneficial to consumers. As the Federal Trade Commission has noted on its website (as of November, 2013): "Most mergers actually benefit competition and consumers by allowing firms to operate more efficiently." At the same time, the FTC recognizes, "Some [mergers] are likely to lessen competition. That, in turn, can lead to higher prices, reduced availability of goods or services, lower quality of products, and less innovation. Indeed, some mergers create a concentrated market, while others enable a single firm to raise prices." The challenge for the antitrust regulators at the FTC and the U.S. Department of Justice is to figure out when a merger may hinder competition. This decision involves both numerical tools and some judgments that are difficult to quantify. The following section helps explain how antitrust laws came about.

What is U.S. Antitrust Law?

In the closing decades of the 1800s, many industries in the U.S. economy were dominated by a single firm that had most of the sales for the entire country. Supporters of these large firms argued that they could take advantage of economies of scale and careful planning to provide consumers with products at low prices. However, critics pointed out that when competition was reduced, these firms were free to charge more and make permanently higher profits, and that without the goading of competition, it was not clear that they were as efficient or innovative as they could be.

In many cases, these large firms were organized in the legal form of a "trust," in which a group of formerly independent firms were consolidated together by mergers and purchases, and a group of "trustees" then ran the companies as if they were a single firm. Thus, when the U.S. government passed the Sherman Antitrust Act in 1890 to limit the power of these trusts, it was called an antitrust law. In an early demonstration of the law's power, the U.S. Supreme Court in 1911 upheld the government's right to break up Standard Oil, which had controlled about 90% of the country's oil refining, into 34 independent firms, including Exxon, Mobil, Amoco, and Chevron. In 1914, theClayton Antitrust Act outlawed mergers and acquisitions (where the outcome would be to "substantially lessen competition" in an industry), price discrimination (where different customers are charged different prices for the same product), and tied sales (where purchase of one product commits the buyer to purchase some other product). Also in 1914, the Federal Trade Commission (FTC) was created to define more specifically what competition was unfair. In 1950, the Celler-Kefauver Act extended the Clayton Act by restricting vertical and conglomerate mergers. In the twenty-first century, the FTC and the U.S. Department of Justice continue to enforce antitrust laws.

The Four-Firm Concentration Ratio

Regulators have struggled for decades to measure the degree of monopoly power in an industry. An early tool was the ***concentration ratio***, which measures what share of the total sales in the industry are accounted for by the largest firms, typically the top four to eight firms. For an explanation of how high market concentrations can create inefficiencies in an economy, refer to Monopoly.

Say that the market for replacing broken automobile windshields in a certain city has 18 firms with the market shares shown in Table 11.1, where the *market share* is each firm's proportion of total sales in that market. The four-firm concentration ratio is calculated by adding the market shares of the four largest firms: in this case, 16 + 10 + 8 + 6 = 40. This concentration ratio would not be considered especially high, because the largest four firms have less than half the market.

Table 11.1 Calculating Concentration Ratios from Market Shares

If the market shares in the market for replacing automobile windshields are:	
Smooth as Glass Repair Company	16% of the market
The Auto Glass Doctor Company	10% of the market
Your Car Shield Company	8% of the market
Seven firms that each have 6% of the market	42% of the market, combined
Eight firms that each have 3% of the market	24% of the market, combined
Then the four-firm concentration ratio is 16 + 10 + 8 + 6 = 40.	

The concentration ratio approach can help to clarify some of the fuzziness over deciding when a merger might affect competition. For instance, if two of the smallest firms in the hypothetical market for repairing automobile windshields merged, the four-firm concentration ratio would not change—which implies that there is not much worry that the degree of competition in the market has notably diminished. However, if the top two firms merged, then the four-firm concentration ratio would become 46 (that is, 26 + 8 + 6 + 6). While this concentration ratio is modestly higher, the four-firm concentration ratio would still be less than half, so such a proposed merger might barely raise an eyebrow among antitrust regulators.

LINK IT UP

Visit this website to read an article about Google's run-in with the FTC.

The Herfindahl-Hirshman Index

A four-firm concentration ratio is a simple tool, which may reveal only part of the story. For example, consider two industries that both have a four-firm concentration ratio of 80. However, in one industry five firms each control 20% of the market, while in the other industry, the top firm holds 77% of the market and all the other firms have 1% each. Although the four-firm concentration ratios are identical, it would be reasonable to worry more about the extent of competition in the second case—where the largest firm is nearly a monopoly—than in the first.

Another approach to measuring industry concentration that can distinguish between these two cases is called the *Herfindahl-Hirschman Index* (*HHI*). The HHI, as it is often called, is calculated by summing the squares of the market share of each firm in the industry.

Calculating HHI

Step 1. Calculate the HHI for a monopoly with a market share of 100%. Because there is only one firm, it has 100% market share. The HHI is 100^2= 10,000.

Step 2. For an extremely competitive industry, with dozens or hundreds of extremely small competitors, the value of the HHI might drop as low as 100 or even less. Calculate the HHI for an industry with 100 firms that each have 1% of the market. In this case, the HHI is $100(1^2) = 100$.

Step 3. Calculate the HHI for the industry shown in Table 11.1. In this case, the HHI is $16^2 + 10^2 + 8^2 + 7(6^2) + 8(3^2) = 744$.

Step 4. Note that the HHI gives greater weight to large firms.

Step 5. Consider the example given earlier, comparing one industry where five firms each have 20% of the market with an industry where one firm has 77% and the other 23 firms have 1% each. The two industries have the same four-firm concentration ratio of 80. But the HHI for the first industry is $5(20^2) = 2{,}000$, while the HHI for the second industry is much higher at $77^2 + 23(1^2) = 5{,}952$.

Step 6. Note that the near-monopolist in the second industry drives up the HHI measure of industrial concentration.

Step 7. Review Table 11.2 which gives some examples of the four-firm concentration ratio and the HHI in various U.S. industries in 2009. (You can find market share data from multiple industry sources. Data in the table are from: Verizon (for wireless), *The Wall Street Journal* (for automobiles), IDC Worldwide (for computers) and the U.S. Bureau of Transportation Statistics (for airlines).)

Table 11.2. Examples of Concentration Ratios and HHIs in the U.S. Economy, 2009

U.S. Industry	Four-Firm Ratio	HHI
Wireless	91	2,311
Largest five: Verizon, AT&T, Sprint Nextel, T-Mobile, MetroPCS		
Automobiles	63	1,121
Largest five: GM, Toyota, Ford, Honda, Chrysler		
Computers	74	1,737
Largest five: HP, Dell, Acer, Apple, Toshiba		
Airlines	44	536
Largest five: Southwest, American, Delta, United, U.S. Airways		

In the 1980s, the FTC followed these guidelines: If a merger would result in an HHI of less than 1,000, the FTC would probably approve it. If a merger would result in an HHI of more than 1,800, the FTC would probably challenge it. If a merger would result in an HHI between 1,000 and 1,800, then the FTC would scrutinize the plan and make a case-by-case decision. However, in the last several decades, the antitrust enforcement authorities have moved away from relying as heavily on measures of concentration ratios and HHIs to determine whether a merger will be allowed, and instead carried out more case-by-case analysis on the extent of competition in different industries.

New Directions for Antitrust

Both the four-firm concentration ratio and the Herfindahl-Hirschman index share some weaknesses. First, they begin from the assumption that the "market" under discussion is well-defined, and the only question is measuring how sales are divided in that market.

Second, they are based on an implicit assumption that competitive conditions across industries are similar enough that a broad measure of concentration in the market is enough to make a decision about the effects of a merger. These assumptions, however, are not always correct. In response to these two problems, the antitrust regulators have been changing their approach in the last decade or two.

Defining a ***market*** is often controversial. For example, Microsoft in the early 2000s had a dominant share of the software for computer operating systems. However, in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 16% in 2000. A narrowly defined market will tend to make concentration appear higher, while a broadly defined market will tend to make it appear smaller.

There are two especially important shifts affecting how markets are defined in recent decades: one centers on technology and the other centers on ***globalization***. In addition, these two shifts are interconnected. With the vast improvement in communications technologies, including the development of the Internet, a consumer can order books or pet supplies from all over the country or the world. As a result, the degree of competition many local retail businesses face has increased. The same effect may operate even more strongly in markets for business supplies, where so-called "business-to-business" websites can allow buyers and suppliers from anywhere in the world to find each other.

Globalization has changed the boundaries of markets. As recently as the 1970s, it was common for measurements of concentration ratios and HHIs to stop at national borders. Now, many industries find that their competition comes from the global market. A few decades ago, three companies, General Motors, Ford, and Chrysler, dominated the U.S. auto market. By 2007, however, these three firms were making less than half of U.S. auto sales, and facing competition from well-known car manufacturers such as Toyota, Honda, Nissan, Volkswagen, Mitsubishi, and Mazda. When HHIs are calculated with a global perspective, concentration in most major industries—including cars—is lower than in a purely domestic context.

Because attempting to define a particular market can be difficult and controversial, the Federal Trade Commission has begun to look less at market share and more at the data on actual competition between businesses. For example, in February 2007, Whole Foods Market and Wild Oats Market announced that they wished to merge. These were the two largest companies in the market that the government defined as “premium natural and organic supermarket chains.” However, one could also argue that they were two relatively small companies in the broader market for all stores that sell groceries or specialty food products.

Rather than relying on a market definition, the government antitrust regulators looked at detailed evidence on profits and prices for specific stores in different cities, both before and after other competitive stores entered or exited. Based on that evidence, the Federal Trade Commission decided to block the merger. After two years of legal battles, the merger was eventually allowed in 2009 under the conditions that Whole Foods sell off the Wild Oats brand name and a number of individual stores, to preserve competition in certain local markets. For more on the difficulties of defining markets, refer to Monopoly.

This new approach to antitrust regulation involves detailed analysis of specific markets and companies, instead of defining a market and counting up total sales. A common starting point is for antitrust regulators to use statistical tools and real-world evidence to estimate the ***demand curves*** and ***supply curves*** faced by the firms that are proposing the merger. A second step is to specify how competition occurs in this specific industry. Some possibilities include competing to cut prices, to raise output, to build a brand name through advertising, and to build a reputation for good service or high quality. With these pieces of the puzzle in place, it is then possible to build a statistical model that estimates the likely outcome for consumers if the two firms are allowed to merge. Of course, these models do require some degree of subjective judgment, and so they can become the subject of legal disputes between the antitrust authorities and the companies that wish to merge.

202. Reading: Regulating Anticompetitive Behavior

Regulating Anticompetitive Behavior

The U.S. antitrust laws reach beyond blocking mergers that would reduce competition to include a wide array of anticompetitive practices. For example, it is illegal for competitors to form a cartel to collude to make pricing and output decisions, as if they were a monopoly firm. The Federal Trade Commission and the U.S. Department of Justice prohibit firms from agreeing to fix prices or output, rigging bids, or sharing or dividing markets by allocating customers, suppliers, territories, or lines of commerce.

In the late 1990s, for example, the antitrust regulators prosecuted an international cartel of vitamin manufacturers, including the Swiss firm Hoffman-La Roche, the German firm BASF, and the French firm Rhone-Poulenc. These firms reached agreements on how much to produce, how much to charge, and which firm would sell to which customers. The high-priced vitamins were then bought by firms like General Mills, Kellogg, Purina-Mills, and Proctor and Gamble, which pushed up the prices more. Hoffman-La Roche pleaded guilty in May 1999 and agreed both to pay a fine of $500 million and to have at least one top executive serve four months of jail time.

Under U.S. antitrust laws, monopoly itself is not illegal. If a firm has a monopoly because of a newly patented invention, for example, the law explicitly allows a firm to earn higher-than-normal profits for a time as a reward for innovation. If a firm achieves a large share of the market by producing a better product at a lower price, such behavior is not prohibited by antitrust law.

Restrictive Practices

Antitrust law includes rules against ***restrictive practices***—practices that do not involve outright agreements to raise price or to reduce the quantity produced, but that might have the effect of reducing competition. Antitrust cases involving restrictive practices are often controversial, because they delve into specific contracts or agreements between firms that are allowed in some cases but not in others.

For example, if a product manufacturer is selling to a group of dealers who then sell to the general public it is illegal for the manufacturer to demand a ***minimum resale price maintenance agreemen***t, which would require the dealers to sell for at least a certain minimum price. A minimum price contract is illegal because it would restrict competition among dealers. However, the manufacturer is legally allowed to "suggest" minimum prices and to stop selling to dealers who regularly undercut the suggested price. If you think this rule sounds like a fairly subtle distinction, you are right.

An ***exclusive dealing*** agreement between a manufacturer and a dealer can be legal or illegal. It is legal if the purpose of the contract is to encourage competition between dealers. For example, it is legal for the Ford Motor Company to sell its cars to only Ford dealers, for General Motors to sell to only GM dealers, and so on. However, exclusive deals may also limit competition. If one large retailer obtained the exclusive rights to be the sole distributor of televisions, computers, and audio equipment made by a number of companies, then this exclusive contract would have an anticompetitive effect on other retailers.

Tie-in sales happen when a customer is allowed to buy one product only if the customer also buys a second product. Tie-in sales may sometimes be illegal. Suppose that to purchase a popular DVD, the store required that you also purchase two other DVDs of your choice. That would be a violation of the tie-in sales regulations.

A related, but not identical, concept is called ***bundling***, where two or more products are sold as one. For example, a cell phone that takes pictures forces a customer to buy a camera along with the phone.

In some cases, tie-ins and bundling can be viewed as anticompetitive. However, in other cases they may be legal and even common. It is common for people to purchase season tickets to a sports team or a set of concerts so that they can be guaranteed tickets to the few contests or shows that are most popular and likely to sell out. Computer software manufacturers may often bundle together a number of different programs, even when the buyer wants only a few of the programs. Think about the software that is included in a new computer purchase, for example.

Predatory pricing occurs when the existing firm (or firms) reacts to a new firm by dropping prices very low, until the new firm is driven out of the market, at which point the existing firm raises prices again. This pattern of pricing is aimed at deterring the entry of new firms into the market. But in practice, it can be hard to figure out when pricing should be considered predatory. Say that American Airlines is flying between two cities, and a new airline starts flying between the same two cities, at a lower price. If American Airlines cuts its price to match the new entrant, is this predatory pricing? Or is it just market competition at work? A commonly proposed rule is that if a firm is selling for less than its average variable cost—that is, at a price where it should be shutting down—then there is evidence for predatory pricing. But calculating in the real world what costs are variable and what costs are fixed is often not obvious, either.

The Microsoft antitrust case embodies many of these gray areas in restrictive practices, as the next section shows.

DID MICROSOFT® ENGAGE IN ANTICOMPETITIVE AND RESTRICTIVE PRACTICES?

The most famous restrictive practices case of recent years was a series of lawsuits by the U.S. government against Microsoft—lawsuits that were encouraged by some of Microsoft's competitors. All sides admitted that Microsoft's Windows program had a near-monopoly position in the market for the software used in general computer operating systems. All sides agreed that the software had many satisfied customers. All sides agreed that the capabilities of computer software that was compatible with Windows—both software produced by Microsoft and that produced by other companies—had expanded dramatically in the 1990s. Having a monopoly or a near-monopoly is not necessarily illegal in and of itself, but in cases where one company controls a great deal of the market, antitrust regulators look at any allegations of restrictive practices with special care.

The antitrust regulators argued that Microsoft had gone beyond profiting from its software innovations and its dominant position in the software market for operating systems, and had tried to use its market power in operating systems software to take over other parts of the software industry. For example, the government argued that Microsoft had engaged in an anticompetitive form of exclusive dealing by threatening computer makers that, if they did not leave another firm's software off their machines (specifically, Netscape's

Internet browser), then Microsoft would not sell them its operating system software. Microsoft was accused by the government antitrust regulators of tying together its Windows operating system software, where it had a monopoly, with its Internet Explorer browser software, where it did not have a monopoly, and thus using this bundling as an anticompetitive tool. Microsoft was also accused of a form of predatory pricing; namely, giving away certain additional software products for free as part of Windows, as a way of driving out the competition from other makers of software.

In April 2000, a federal court held that Microsoft's behavior had crossed the line into unfair competition, and recommended that the company be broken into two competing firms. However, that penalty was overturned on appeal, and in November 2002 Microsoft reached a settlement with the government that it would end its restrictive practices.

The concept of restrictive practices is continually evolving, as firms seek new ways to earn profits and government regulators define what is permissible and what is not. A situation where the law is evolving and changing is always somewhat troublesome, since laws are most useful and fair when firms know what they are in advance. In addition, since the law is open to interpretation, competitors who are losing out in the market can accuse successful firms of anticompetitive restrictive practices, and try to win through government regulation what they have failed to accomplish in the market. Officials at the Federal Trade Commission and the Department of Justice are, of course, aware of these issues, but there is no easy way to resolve them.

203. Reading: Regulating Natural Monopolies

Regulating Natural Monopolies

Most true monopolies today in the U.S. are regulated, natural monopolies. A natural monopoly poses a difficult challenge for competition policy, because the structure of costs and demand seems to make competition unlikely or costly. A ***natural monopoly*** arises when average costs are declining over the range of production that satisfies market demand. This typically happens when fixed costs are large relative to variable costs. As a result, one firm is able to supply the total quantity demanded in the market at lower cost than two or more firms—so splitting up the natural monopoly would raise the average cost of production and force customers to pay more.

Public utilities, the companies that have traditionally provided water and electrical service across much of the United States, are leading examples of natural monopoly. It would make little sense to argue that a local water company should be broken up into several competing companies, each with its own separate set of pipes and water supplies. Installing four or five identical sets of pipes under a city, one for each water company, so that each household could choose its own water provider, would be terribly costly. The same argument applies to the idea of having many competing companies for delivering electricity to homes, each with its own set of wires. Before the advent of wireless phones, the argument also applied to the idea of many different phone companies, each with its own set of phone wires running through the neighborhood.

THE CHOICES IN REGULATING A NATURAL MONOPOLY

So what then is the appropriate competition policy for a natural monopoly? Figure 11.3 illustrates the case of natural monopoly, with a market demand curve that cuts through the downward-sloping portion of the ***average cost curve***. Points A, B, C, and F illustrate four of the main choices for regulation. Table 11.3 outlines the regulatory choices for dealing with a natural monopoly.

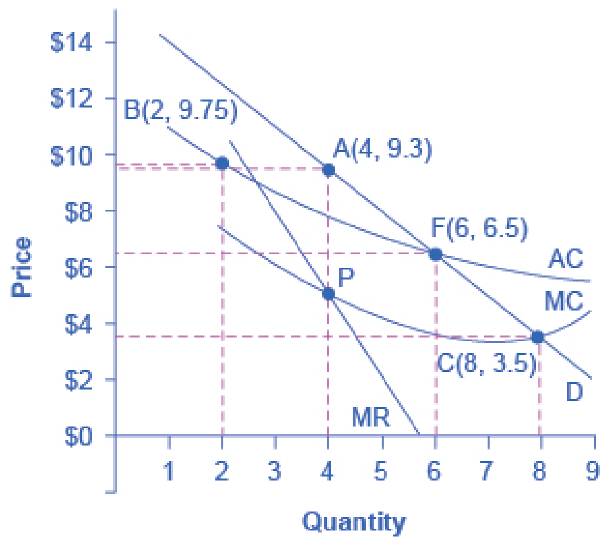


Figure 11.3. *Regulatory Choices in Dealing with Natural Monopoly. A natural monopoly will maximize profits by producing at the quantity where marginal revenue (MR) equals marginal costs (MC) and by then looking to the market demand curve to see what price to charge for this quantity. This monopoly will produce at point A, with a quantity of 4 and a price of 9.3. If antitrust regulators split this company exactly in half, then each half would produce at point B, with average costs of 9.75 and output of 2. The regulators might require the firm to produce where marginal cost crosses the market demand curve at point C. However, if the firm is required to produce at a quantity of 8 and sell at a price of 3.5, the firm will suffer from losses. The most likely choice is point F, where the firm is required to produce a quantity of 6 and charge a price of 6.5.*

Table 11.3 Regulatory Choices in Dealing with Natural Monopoly

Quantity	Price	Total Revenue*	Marginal Revenue	Total Cost	Marginal Cost	Average Cost
1	14.7	14.7	–	11.0	–	11.00
2	12.4	24.7	10.0	19.5	8.5	9.75
3	10.6	31.7	7.0	25.5	6.0	8.50
4	9.3	37.2	5.5	31.0	5.5	7.75
5	8.0	40.0	2.8	35.0	4.0	7.00
6	6.5	39.0	−1.0	39.0	4.0	6.50
7	5.0	35.0	−4.0	42.0	3.0	6.00
8	3.5	28.0	−7.0	45.5	3.5	5.70
9	2.0	18.0	−10.0	49.5	4.0	5.5

*Total Revenue is given by multiplying price and quantity. However, some of the price values in this table have been rounded for ease of presentation.

The first possibility is to leave the natural monopoly alone. In this case, the monopoly will follow its normal approach to maximizing profits. It determines the quantity where MR = MC, which happens at point P at a quantity of 4. The firm then looks to point A on the demand curve to find that it can charge a price of 9.3 for that profit-maximizing quantity. Since the price is above the average cost curve, the natural monopoly would earn economic profits.

A second outcome arises if antitrust authorities decide to divide the company, so that the new firms can compete. As a simple example, imagine that the company is cut in half. Thus, instead of one large firm producing a quantity of 4, two half-size firms each produce a quantity of 2. Because of the declining average cost curve (AC), the average cost of production for each of the half-size companies each producing 2, as shown at point B, would be 9.75, while the average cost of production for a larger firm producing 4 would only be 7.75. Thus, the economy would become less productively efficient, since the good is being produced at a higher average cost. In a situation with a downward-sloping average cost curve, two smaller firms will always have higher average costs of production than one larger firm for any quantity of total output. In addition, the antitrust authorities must worry that splitting the natural monopoly into pieces may be only the start of their problems. If one of the two firms grows larger than the other, it will have lower average costs and may be able to drive its competitor out of the market. Alternatively, two firms in a market may discover subtle ways of coordinating their behavior and keeping prices high. Either way, the result will not be the greater competition that was desired.

A third alternative is that regulators may decide to set prices and quantities produced for this industry. The regulators will try to choose a point along the market demand curve that benefits both consumers and the broader social interest. Point C illustrates one tempting choice: the regulator requires that the firm produce the quantity of output where marginal cost crosses the demand curve at an output of 8, and charge the price of 3.5, which is equal to ***marginal cost*** at that point. This rule is appealing because it requires price to be set equal to marginal cost, which is what would occur in a perfectly competitive market, and it would assure consumers a higher quantity and lower price than at the monopoly choice A. In fact, efficient allocation of resources would occur at point C, since the value to the consumers of the last unit bought and sold in this market is equal to the marginal cost of producing it.

Attempting to bring about point C through force of regulation, however, runs into a severe difficulty. At point C, with an output of 8, a price of 3.5 is below the average cost of production, which is 5.7, and so if the firm charges a price of 3.5, it will be suffering losses. Unless the regulators or the government offer the firm an ongoing public subsidy (and there are numerous political problems with that option), the firm will lose money and go out of business.

Perhaps the most plausible option for the regulator is point F; that is, to set the price where AC crosses the demand curve at an output of 6 and a price of 6.5. This plan makes some sense at an intuitive level: let the natural monopoly charge enough to cover its average costs and earn a normal rate of profit, so that it can continue operating, but prevent the firm from raising prices and earning abnormally high monopoly profits, as it would at the monopoly choice A. Of course, determining this level of output and price with the political pressures, time constraints, and limited information of the real world is much harder than identifying the point on a graph. For more on the problems that can arise from a centrally determined price, see the discussion of price floors and price ceilings in Demand and Supply.

COST-PLUS VERSUS PRICE CAP REGULATION

Indeed, regulators of public utilities for many decades followed the general approach of attempting to choose a point like F in Figure 11.3. They calculated the average cost of production for the water or electricity companies, added in an amount for the normal rate of profit the firm should expect to earn, and set the price for consumers accordingly. This method was known as ***cost-plus regulation***.

Cost-plus regulation raises difficulties of its own. If producers are reimbursed for their costs, plus a bit more, then at a minimum, producers have less reason to be concerned with high costs—because they can just pass them along in higher prices. Worse, firms under cost-plus regulation even have an incentive to generate high costs by building huge factories or employing lots of staff, because what they can charge is linked to the costs they incur.

Thus, in the 1980s and 1990s, some regulators of public utilities began to use ***price cap regulation***, where the regulator sets a price that the firm can charge over the next few years. A common pattern was to require a price that declined slightly over time. If the firm can find ways of reducing its costs more quickly than the price caps, it can make a high level of profits. However, if the firm cannot keep up with the price caps or suffers bad luck in the market, it may suffer losses. A few years down the road, the regulators will then set a new series of price caps based on the firm's performance.

Price cap regulation requires delicacy. It will not work if the price regulators set the price cap unrealistically low. It may not work if the market changes dramatically so that the firm is doomed to incurring losses no matter what it does—say, if energy prices rise dramatically on world markets, then the company selling natural gas or heating oil to homes may not be able to meet price caps that seemed reasonable a year or two ago. But if the regulators compare the prices with producers of the same good in other areas, they can, in effect, pressure a natural monopoly in one area to compete with the prices being charged in other areas. Moreover, the possibility of earning greater profits or experiencing losses—instead of having an average rate of profit locked in every year by cost-plus regulation—can provide the natural monopoly with incentives for efficiency and innovation.

With natural monopoly, market competition is unlikely to take root, so if consumers are not to suffer the high prices and restricted output of an unrestricted monopoly, government regulation will need to play a role. In attempting to design a system of price cap regulation with flexibility and incentive, government regulators do not have an easy task.

204. Reading: The Great Deregulation Experiment

Doubts about Regulation of Prices and Quantities

Governments at all levels across the United States have regulated prices in a wide range of industries. In some cases, like water and electricity that have natural monopoly characteristics, there is some room in economic theory for such regulation. But once politicians are given a basis to intervene in markets and to choose prices and quantities, it is hard to know where to stop.

Beginning in the 1970s, it became clear to policymakers of all political leanings that the existing price regulation was not working well. The United States carried out a great policy experiment—the ***deregulation*** discussed in Monopoly—removing government controls over prices and quantities produced in airlines, railroads, trucking, intercity bus travel, natural gas, and bank interest rates. The next section discusses the outcome of deregulation in one industry in particular—airlines.

WHAT ARE THE RESULTS OF AIRLINE

DEREGULATION?

Why did the pendulum swing in favor of deregulation? Consider the airline industry. In the early days of air travel, no airline could make a profit just by flying passengers. Airlines needed something else to carry and the Postal Service provided that something with airmail. And so the first U.S. government regulation of the airline industry happened through the Postal Service, when in 1926 the Postmaster General began giving airlines permission to fly certain routes based on the needs of mail delivery—and the airlines took some passengers along for the ride. In 1934, the Postmaster General was charged by the antitrust authorities with colluding with the major airlines of that day to monopolize the nation's airways. In 1938, the Civil Aeronautics Board (CAB) was created to regulate airfares and routes instead. For 40 years, from 1938 to 1978, the CAB approved all fares, controlled all entry and exit, and specified which airlines could fly which routes. There was zero entry of new airlines on the main routes across the country for 40 years, because the CAB did not think it was necessary.

In 1978, the Airline Deregulation Act took the government out of the business of determining airfares and schedules. The new law shook up the industry. Famous old airlines like Pan American, Eastern, and Braniff went bankrupt and disappeared. Some new airlines like People Express were created—and then vanished.

The greater competition from deregulation reduced airfares by about one-third over the next two decades, saving consumers billions of dollars a year. The average flight used to take off with just half its seats full; now it is two-thirds full, which is far more efficient. Airlines have also developed hub-and-spoke systems, where planes all fly into a central hub city at a certain time and then depart. As a result, one can fly between any of the spoke cities with just one connection—and there is greater service to more cities than before deregulation. With lower fares and more service, the number of air passengers doubled from the late 1970s to the start of the 2000s—an increase that, in turn, doubled the number of jobs in the airline industry. Meanwhile, with the watchful oversight of government safety inspectors, commercial air travel has continued to get safer over time.

The U.S. airline industry is far from perfect. For example, a string of mergers in recent years has raised concerns over how competition might be compromised.

One difficulty with government price regulation is what economists call ***regulatory capture***, in which the firms supposedly being regulated end up playing a large role in setting the regulations that they will follow. When the airline industry was being regulated, for example, it suggested appointees to the regulatory board, sent lobbyists to argue with the board, provided most of the information on which the board made decisions, and offered well-paid jobs to at least some of the people leaving the board. In this situation, consumers can easily end up being not very well represented by the regulators. The result of regulatory capture is that government price regulation can often become a way for existing competitors to work together to reduce output, keep prices high, and limit competition.

The Effects of Deregulation

Deregulation, both of airlines and of other industries, has its negatives. The greater pressure of competition led to entry and exit. When firms went bankrupt or contracted substantially in size, they laid off workers who had to find other jobs. Market competition is, after all, a full-contact sport.

A number of major accounting scandals involving prominent corporations such as Enron, Tyco International, and WorldCom led to the ***Sarbanes-Oxley Act*** in 2002. Sarbanes-Oxley was designed to increase confidence in financial information provided by public corporations to protect investors from accounting fraud.

The Great Recession which began in late 2007 and which the U.S. economy is still struggling to recover from was caused at least in part by a global financial crisis, which began in the United States. The key component of the crisis was the creation and subsequent failure of several types of unregulated financial assets, such as collateralized mortgage obligations (CMOs, a type of mortgage-backed security), and credit default swaps (CDSs, insurance contracts on assets like CMOs that provided a payoff even if the holder of the CDS did not own the CMO). Many of these assets were rated very safe by private credit rating agencies such as Standard & Poors, Moody's, and Fitch.

The collapse of the markets for these assets precipitated the financial crisis and led to the failure of Lehman Brothers, a major investment bank, numerous large commercial banks, such as Wachovia, and even the Federal National Mortgage Corporation (Fannie Mae), which had to be nationalized—that is, taken over by the federal government. One response to the financial crisis was the *Dodd-Frank Act*, which attempted major reforms of the financial system. The legislation's purpose, as noted on dodd-frank.com is:

> To promote the financial stability of the United States by improving accountability and transparency in the financial system, to end "too big to fail," to protect the American taxpayer by ending bailouts, [and] to protect consumers from abusive financial services practices. . .

We will explore the financial crisis and the Great Recession in more detail in the macroeconomic modules of this text, but for now it should be clear that many Americans have grown disenchanted with deregulation, at least of financial markets.

All market-based economies operate against a background of laws and regulations, including laws about enforcing contracts, collecting taxes, and protecting health and the environment. The government policies discussed in this module—like blocking certain anticompetitive mergers, ending restrictive practices, imposing price cap regulation on natural monopolies, and deregulation—demonstrate the role of government to strengthen the incentives that come with a greater degree of competition.

MORE THAN COOKING, HEATING, AND COOLING

What did the Federal Trade Commission (FTC) decide on the Kinder Morgan/El Paso Corporation merger? After careful examination, federal officials decided there was only one area of significant overlap that might provide the merged firm with strong market power. The FTC approved the merger, provided Kinder Morgan divest itself of the overlap area. Tallgrass purchased Kinder Morgan Interstate Gas Transmission, Trailblazer Pipeline Co. LLC, two processing facilities in Wyoming, and Kinder Morgan's 50 percent interest in the Rockies Express Pipeline to meet the FTC requirements. The FTC was attempting to strike a balance between potential cost reductions resulting from economies of scale and concentration of market power.

Did the price of natural gas decrease? Yes, rather significantly. In 2010, the wellhead price of natural gas was $4.48 per thousand cubic foot; in 2012 the price had fallen to just $2.66. Was the merger responsible for the large drop in price? The answer is uncertain. The larger contributor to the sharp drop in price was the overall increase in the supply of natural gas. More and more natural gas was able to be recovered by fracturing shale deposits, a process called fracking. Fracking, which is controversial for environmental reasons, enabled the recovery of known reserves of natural gas that previously were not economically feasible to tap. Kinder Morgan's control of 80,000-plus miles of pipeline likely made moving the gas from wellheads to end users smoother and allowed for an even greater benefit from the increased supply.

Self Check: Controlling Monopolies

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the five Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/672

205. Putting It Together: Monopoly

Summary

The goal of this module was analyze a firm's profit maximizing strategies under conditions of monopoly. You learned how to:

- **Define the characteristics of a monopoly.** A monopoly occurs when a single firm supplies the whole market for some product. Because they face no direct competition, monopolies can charge any price they want and earn economic profits, even in the long run.
- **Define and explain the sources of barriers to entry.** Barriers to entry are economic or legal prohibitions on other firms entering an industry to capture some of the monopoly's profits.
- **Calculate and graph a monopoly's fixed, variable, average, marginal and total costs.** Costs are computed and cost curves graphed the same way as in perfect competition. This is one of the similarities across market structures.
- **Explain why a monopoly is inefficient using deadweight loss.** Allocative inefficiency occurs when firms produce less than the optimal supply, which monopolies do to allow them to charge a higher price. Deadweight loss is the loss in total surplus (producer + consumer surplus) that occurs at output less than the optimal one.
- **Analyze different strategies to control monopolies, including natural monopolies.** Governments use laws and regulation to reduce the inefficiency of monopolies. Regulated monopolies supply more than they otherwise would in return for a

guaranteed profit.

Examples

Why are monopolies bad? It's not primarily because they charge too high a price. Rather, it's because they are allocatively inefficient, in other words, they produce too little of the product, and because (often) they are productively inefficient, that is, they don't produce as cheaply as possible. That last point is a bit subtle, so we'll learn more about it in a later module.

New Packaging Display by tales of a wandering youkai. CC-BY.

Nearly all monopolies in the U.S. are regulated monopolies, meaning the prices they charge have to be reviewed and approved (or not) by a regulatory branch of the government. A good example of this is the U.S. Postal Service, whose rates must be accepted by the Postal Rate Commission. Thus, regulated monopolies don't behave exactly the way pure monopolies do, as explained in this module. Why do we study them then? Because nearly all firms in the real world have some market power, that is the ability to influence the market price. Monopoly power, the ability to set the market price, is the ultimate in market power. Understanding how monopolies exploit their power helps us understand how real world, but not-quite-monopoly firms, operate.

206. Glossary: Monopoly

allocative efficiency
: producing the optimal quantity of some output; the quantity where the marginal benefit to society of one more unit just equals the marginal cost

barriers to entry
: the legal, technological, or market forces that may discourage or prevent potential competitors from entering a market

copyright
: a form of legal protection to prevent copying, for commercial purposes, original works of authorship, including books and music

deregulation
: removing government controls over setting prices and quantities in certain industries

intellectual property
: the body of law including patents, trademarks, copyrights, and trade secret law that protect the right of inventors to produce and sell their inventions

legal monopoly
: legal prohibitions against competition, such as regulated monopolies and intellectual property protection

marginal profit
: profit of one more unit of output, computed as marginal revenue minus marginal cost

monopoly
: a situation in which one firm produces all of the output in a market

natural monopoly
economic conditions in the industry, for example, economies of scale or control of a critical resource, that limit effective competition

patent
a government rule that gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time

predatory pricing
when an existing firm uses sharp but temporary price cuts to discourage new competition

trade secrets
methods of production kept secret by the producing firm

trademark
an identifying symbol or name for a particular good and can only be used by the firm that registered that trademark

207. Discussion: Price Discrimination

Draw the graph for a monopoly with demand, marginal revenue, and marginal cost curves. Identify the profit-maximizing output level (Qm) and price (Pm).

Suppose the monopolist sells Qm units of output at the regular price and then puts the product on sale at a lower price, Ps. Show the new price and quantity. Identify the consumer surplus of the additional sales. What happens to the firm's profits? Does price discrimination lead to a more efficient or less efficient outcome? Why or why not?

PART XII

MODULE: MONOPOLISTICALLY COMPETITIVE INDUSTRY

208. Why It Matters: Monopolistically Competitive Industries

Why analyze a firm's profit maximizing strategies under conditions of monopolistic competition?

This module explains monopolistic competition, the second example of imperfect, or real world, competition (along with oligopoly, which you studied in the previous module). Most of what you purchase at the retail level is from monopolistically competitive firms, so this model is relevant to most people's lives.

Fast Food is the Best! by ebru, CC-BY.

Monopolistically competitive industries are those that contain more than a few firms, each of which offers a similar but not identical product. Take fast food, for example. The fast food market is quite competitive, and yet each firm has a monopoly in its own product. Some customers have a preference for McDonald's over Burger King. Some have a preference for Dominoes over Pizza Hut. These preferences give monopolistically competitive firms market power, which they can exploit to earn positive economic profits.

Consider the following questions:

- Why do gas stations charge different prices for a gallon of gasoline?
- What determines how far apart the prices of Colgate and Crest

toothpaste can be?

- Why did fast food restaurants start offering salads?
- Why are fast food chicken sandwich prices different from burger prices?
- Why did McDonalds come up with the Big Mac sandwich?

In a real sense, the model of monopolistic competition is a combination of the models of perfect competition and monopoly. As you progress through this module, think about the similarities and the differences between monopolistic competition and those other two models of market structure.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=237

LEARNING OUTCOMES

- Define the characteristics of a monopolistically competitive industry
- Calculate and graph the firm's fixed, variable, average, marginal and total costs
- Explain the difference between short run and long run equilibrium in a monopolistically competitive industry
- Understand how product differentiation works in monopolistically competitive industries and how firms use advertising to differentiate their products, understanding impact on elasticity
- Understand why monopolistically competitive markets are

inefficient (including deadweight loss)

209. Outcome: Monopolistically Competitive Industries

What you'll learn to do: define the characteristics of a monopolistically competitive industry

In this outcome, you will come to understand how and why some markets are NOT perfectly competitive, but more closely resemble markets for monopolies.

Here are some of the specific things you'll learn to do in this section:

- Understand the difference between the firm and the industry

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Introducing Monopolistic Competition
- Reading: Monopolistic Competition
- Self Check: Monopolistically Competitive Industries

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

210. Reading: Introducing Monopolistic Competition

THE TEMPTATION TO DEFY THE LAW

Figure 10.1. Competing Brands? The laundry detergent market is one that is characterized neither as perfect competition nor monopoly. (Credit: modification of work by Pixel Drip/ Flickr Creative Commons)

Laundry detergent and bags of ice—products of industries that seem pretty mundane, maybe even boring. Hardly! Both have been the center of clandestine meetings and secret deals worthy of a spy novel. In France, between 1997 and 2004, the top four laundry detergent producers (Proctor & Gamble, Henkel, Unilever, and Colgate-Palmolive) controlled about 90 percent of the French soap market. Officials from the soap firms were meeting secretly, in out-of-the-way, small cafés around Paris. Their goals: Stamp out competition and set prices.

Around the same time, the top five Midwest ice makers (Home City Ice, Lang Ice, Tinley Ice, Sisler's Dairy, and Products of Ohio) had similar goals in mind when they secretly agreed to divide up the bagged ice market.

If both groups could meet their goals, it would enable each to act as though they were a single firm—in essence, a monopoly—and enjoy monopoly-size profits. The problem? In many parts of the world, including the European Union and the United States, it is illegal for firms to divide up markets and set prices collaboratively.

These two cases provide examples of markets that are characterized neither as perfect competition nor monopoly.

Instead, these firms are competing in market structures that lie between the extremes of monopoly and perfect competition. How do they behave? Why do they exist?

Introduction to Monopolistic Competition and Oligopoly

Perfect competition and monopoly are at opposite ends of the competition spectrum. A perfectly competitive market has many firms selling identical products, who all act as price takers in the face of the competition. If you recall, ***price takers*** are firms that have no market power. They simply have to take the market price as given.

Monopoly arises when a single firm sells a product for which there are no close substitutes. Microsoft, for instance, has been considered a monopoly because of its domination of the operating systems market.

What about the vast majority of real world firms and organizations that fall between these extremes, firms that could be described as ***imperfectly competitive***? What determines their behavior? They have more influence over the price they charge than perfectly competitive firms, but not as much as a monopoly would. What will they do?

One type of imperfectly competitive market is called ***monopolistic competition***. Monopolistically competitive markets feature a large number of competing firms, but the products that they sell are not identical. Consider, as an example, the Mall of America in Minnesota, the largest shopping mall in the United States. In 2010, the Mall of America had 24 stores that sold women's "ready-to-wear" clothing (like Ann Taylor and Coldwater Creek), another 50 stores that sold clothing for both men and women (like Banana Republic, J. Crew, and Nordstrom's), plus 14 more stores that sold women's specialty clothing (like Motherhood Maternity and Victoria's Secret). Most of the markets that consumers encounter at the retail level are monopolistically competitive.

211. Reading: Monopolistic Competition

Monopolistic Competition

Monopolistic competition involves many ***firms*** competing against each other, but selling products that are distinctive in some way. Examples include stores that sell different styles of clothing; restaurants or grocery stores that sell different kinds of food; and even products like golf balls or beer that may be at least somewhat similar but differ in public perception because of advertising and brand names. When products are distinctive, each firm has a mini-monopoly on its particular style or flavor or brand name. However, firms producing such products must also compete with other styles and flavors and brand names. The term “monopolistic competition” captures this mixture of mini-monopoly and tough competition.

Who invented the theory of imperfect competition?

The theory of imperfect competition was developed by two economists independently but simultaneously in 1933. The first was Edward Chamberlin of Harvard University who published *The Economics of Monopolistic Competition*. The second was Joan Robinson of Cambridge University who published *The Economics of Imperfect Competition*. Robinson subsequently became interested in macroeconomics where she became a prominent Keynesian, and later a post-Keynesian economist. (See The Keynesian Perspective modules for more on Keynes.)

DIFFERENTIATED PRODUCTS

A firm can try to make its products different from those of its competitors in several ways: physical aspects of the product, location from which the product is sold, intangible aspects of the product, and perceptions of the product. Products that are distinctive in one of these ways are called ***differentiated products***.

Physical aspects of a product include all the phrases you hear in advertisements: unbreakable bottle, nonstick surface, freezer-to-microwave, non-shrink, extra spicy, newly redesigned for your comfort. The location of a firm can also create a difference between producers. For example, a gas station located at a heavily traveled intersection can probably sell more gas, because more cars drive by that corner. A supplier to an automobile manufacturer may find that it is an advantage to locate close to the car factory.

Intangible aspects can differentiate a product, too. Some intangible aspects may be promises like a guarantee of satisfaction or money back, a reputation for high quality, services like free delivery, or offering a loan to purchase the product. Finally, ***product differentiation*** may occur in the minds of buyers. For example, many people could not tell the difference in taste between common varieties of beer or cigarettes if they were blindfolded but, because of past habits and advertising, they have strong preferences for certain brands. Advertising can play a role in shaping these intangible preferences.

The concept of differentiated products is closely related to the degree of variety that is available. If everyone in the economy wore only blue jeans, ate only white bread, and drank only tap water, then the markets for clothing, food, and drink would be much closer to perfectly competitive. The variety of styles, flavors, locations, and characteristics creates product differentiation and monopolistic competition.

PERCEIVED DEMAND FOR A MONOPOLISTIC COMPETITOR

A monopolistically competitive firm perceives a demand for its goods that is an intermediate case between monopoly and competition. Figure 10.2 offers a reminder that the ***demand curve*** as faced by a perfectly competitive firm is ***perfectly elastic*** or flat, because the perfectly competitive firm can sell any quantity it wishes at the prevailing ***market price***. In contrast, the demand curve, as faced by a monopolist, is the market demand curve, since a monopolist is the only firm in the market, and hence is downward sloping.

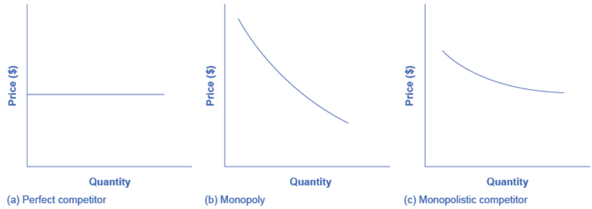


Figure 10.2. *Perceived Demand for Firms in Different Competitive Settings The demand curve faced by a perfectly competitive firm is perfectly elastic, meaning it can sell all the output it wishes at the prevailing market price. The demand curve faced by a monopoly is the market demand. It can sell more output only by decreasing the price it charges. The demand curve faced by a monopolistic ally competitive firm falls in between.*

The demand curve as faced by a monopolistic competitor is not flat, but rather downward-sloping, which means that the monopolistic competitor can raise its price without losing all of its customers or lower the price and gain more customers. Since there are substitutes, the demand curve facing a monopolistically competitive firm is more elastic than that of a monopoly where there are no close substitutes. If a monopolist raises its price, some consumers will choose not to purchase its product—but they will then need to buy a completely different product. However, when a monopolistic competitor raises its price, some consumers will choose not to purchase the product at all, but others will choose to buy a similar product from another firm. If a monopolistic competitor raises its price, it will not lose as many customers as would a perfectly competitive firm, but it will lose more customers than would a monopoly that raised its prices.

At a glance, the demand curves faced by a monopoly and by a monopolistic competitor look similar—that is, they both slope down. But the underlying economic meaning of these perceived demand curves is different, because a monopolist faces the market demand curve and a monopolistic competitor does not. Rather, a monopolistically competitive firm's demand curve is but one of many firms that make up the "before" market demand curve. Are you following? If so, how would you categorize the market for golf balls?

ARE GOLF BALLS REALLY DIFFERENTIATED PRODUCTS?

Monopolistic competition refers to an industry that has more than a few firms, each offering a product which, from the consumer's perspective, is different from its competitors. The U.S. Golf Association runs a laboratory that tests 20,000 golf balls a year. There are strict rules for what makes a golf ball legal. The weight of a golf ball cannot exceed 1.620 ounces and its diameter cannot be less than 1.680 inches (which is a weight of 45.93 grams and a diameter of 42.67 millimeters, in case you were wondering). The balls are also tested by being hit at different speeds. For example, the distance test involves having a mechanical golfer hit the ball with a titanium driver and a swing speed of 120 miles per hour. As the testing center explains: "The USGA system then uses an array of sensors that accurately measure the flight of a golf ball during a short, indoor trajectory from a ball launcher. From this flight data, a computer calculates the lift and drag forces that are generated by the speed, spin, and dimple pattern of the ball. … The distance limit is 317 yards."

Over 1800 golf balls made by more than 100 companies meet the USGA standards. The balls do differ in various ways, like the pattern of dimples on the ball, the types of plastic used on the cover and in the cores, and so on. Since all balls need to conform to the USGA tests, they are much more alike than different. In other words, golf ball manufacturers are monopolistically competitive.

However, retail sales of golf balls are about $500 million per year, which means that a lot of large companies have a powerful incentive to persuade players that golf balls are highly differentiated and that it makes a huge difference which one you choose. Sure, Tiger Woods can tell the difference. For the average duffer (golf-speak for a "mediocre player") who plays a few times a summer—and who loses a lot of golf balls to the woods and lake and needs to buy new ones—most golf balls are pretty much indistinguishable.

Self Check: Monopolistically Competitive Industries

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/677

212. Outcome: Calculating Cost and Revenue

What you'll learn to do: calculate and graph the firm's fixed, variable, average, marginal and total costs

In this section, you will learn how to analyze the cost and revenue curves related to monopolistically competitive firms and use these graphs to determine the best price and quantity for a firm's product.

Here are some of the specific things you'll learn to do in this section:

- Measure variable and total costs as the area under the average variable and average total cost curves
- Calculate and graph the firm's average, marginal and total revenues;
- Measure total revenues as the area under the average revenue curves
- Determine the profit maximizing output level and price; is able to calculate and graphically illustrate where marginal revenue equals marginal costs
- Calculate and graphically illustrate profit and losses for a monopolistically competitive firm

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Choosing Price and Quantity
- Self Check: Calculating Cost, Revenue, Profit and Losses

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

213. Reading: Choosing Price and Quantity

Choosing Price and Quantity

The monopolistically competitive firm decides on its profit-maximizing quantity and price in much the same way as a monopolist. A monopolistic competitor, like a monopolist, faces a downward-sloping demand curve, and so it will choose some combination of price and quantity along its perceived demand curve.

As an example of a ***profit-maximizing*** monopolistic competitor, consider the Authentic Chinese Pizza store, which serves pizza with cheese, sweet and sour sauce, and your choice of vegetables and meats. Although Authentic Chinese Pizza must compete against other pizza businesses and restaurants, it has a differentiated product. The firm's perceived demand curve is downward sloping, as shown in Figure 10.3 and the first two columns of Table 10.1.

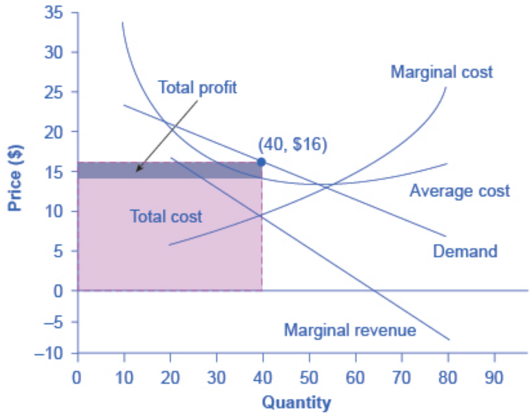


Figure 10.3. *How a Monopolistic Competitor Chooses its Profit Maximizing Output and Price To maximize profits, the Authentic Chinese Pizza shop would choose a quantity where marginal revenue equals marginal cost, or Q where MR = MC. Here it would choose a quantity of 40 and a price of $16.*

Table 10.1. Revenue and Cost Schedule

Quantity	Price	Total Revenue	Marginal Revenue	Total Cost	Marginal Cost	Average Cost
10	$23	$230	—	$340	–	$34
20	$20	$400	$17	$400	$6	$20
30	$18	$540	$14	$480	$8	$16
40	$16	$640	$10	$580	$10	$14.50
50	$14	$700	$6	$700	$12	$14
60	$12	$720	$2	$840	$14	$14
70	$10	$700	–$2	$1,020	$18	$14.57
80	$8	$640	–$6	$1,280	$26	$16

The combinations of price and quantity at each point on the demand curve can be multiplied to calculate the total revenue that the firm would receive, which is shown in the third column of Table 10.1. The fourth column, marginal revenue, is calculated as the change in total revenue divided by the change in quantity. The final columns of Table 10.1 show total cost, marginal cost, and average cost. As always, marginal cost is calculated by dividing the change in total cost by the change in quantity, while average cost is calculated by dividing total cost by quantity. The following example shows how these firms calculate how much of its product to supply at what price.

How a Monopolistic Competitor Determines How Much to Produce and at What Price

The process by which a monopolistic competitor chooses its profit-maximizing quantity and price resembles closely how a monopoly makes these decisions process. First, the firm selects the profit-maximizing quantity to produce. Then the firm decides what price to charge for that quantity.

Step 1. The monopolistic competitor determines its profit-maximizing level of output. In this case, the Authentic Chinese Pizza company will determine the profit-maximizing quantity to produce by considering its marginal revenues and marginal costs. Two scenarios are possible:

- If the firm is producing at a quantity of output where marginal revenue exceeds marginal cost, then the firm should keep expanding production, because each marginal unit is adding to profit by bringing in more revenue than its cost. In this way, the firm will produce up to the quantity where MR = MC.
- If the firm is producing at a quantity where marginal costs exceed marginal revenue, then each marginal unit is costing more than the revenue it brings in, and the firm will increase its profits by reducing the quantity of output until MR = MC.

In this example, MR and MC intersect at a quantity of 40, which is the profit-maximizing level of output for the firm.

Step 2. The monopolistic competitor decides what price to charge. When the firm has determined its profit-maximizing quantity of output, it can then look to its perceived demand curve to find out what it can charge for that quantity of output. On the graph, this process can be shown as a vertical line reaching up through the profit-maximizing quantity until it hits the firm's perceived demand curve. For Authentic Chinese Pizza, it should charge a price of $16 per pizza for a quantity of 40.

Once the firm has chosen price and quantity, it's in a position to calculate total revenue, total cost, and profit. At a quantity of 40, the price of $16 lies above the average cost curve, so the firm is making economic profits. From Table 10.1 we can see that, at an output of 40, the firm's total revenue is $640 and its total cost is $580, so profits are $60. In Figure 10.3, the firm's total revenues are the rectangle with the quantity of 40 on the horizontal axis and the price of $16 on the vertical axis. The firm's total costs are the light shaded rectangle with the same quantity of 40 on the horizontal axis but the average cost of $14.50 on the vertical axis. Profits are total revenues minus total costs, which is the shaded area above the average cost curve.

Although the process by which a monopolistic competitor makes decisions about quantity and price is similar to the way in which a monopolist makes such decisions, two differences are worth remembering. First, although both a monopolist and a monopolistic competitor face downward-sloping demand curves, the monopolist's perceived demand curve is the market demand curve, while the perceived *demand curve* for a monopolistic competitor is based on the extent of its product differentiation and how many competitors it faces. Second, a monopolist is surrounded by barriers to entry and need not fear entry, but a monopolistic competitor who earns profits must expect the entry of firms with similar, but differentiated, products.

Self Check: Calculating Cost, Revenue, Profit and Losses

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/771

214. Outcome: Short Run and Long Run Equilibrium

What you'll learn to do: explain the difference between short run and long run equilibrium in a monopolistically competitive industry

When others notice a monopolistically competitive firm making profits, they will want to enter the market. These new firms entering the market will drive the economic profits towards zero in the long-run.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Monopolistic Competitors and Entry
- Self Check: Short Run and Long Run Equilibrium

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

215. Reading: Monopolistic Competitors and Entry

Monopolistic Competitors and Entry

If one monopolistic competitor earns positive economic profits, other firms will be tempted to enter the market. A gas station with a great location must worry that other gas stations might open across the street or down the road—and perhaps the new gas stations will sell coffee or have a carwash or some other attraction to lure customers. A successful restaurant with a unique barbecue sauce must be concerned that other restaurants will try to copy the sauce or offer their own unique recipes. A laundry detergent with a great reputation for quality must be concerned that other competitors may seek to build their own reputations.

The entry of other firms into the same general market (like gas, restaurants, or detergent) shifts the demand curve faced by a monopolistically competitive firm. As more firms enter the market, the quantity demanded at a given price for any particular firm will decline, and the firm's perceived demand curve will shift to the left. As a firm's perceived demand curve shifts to the left, its marginal revenue curve will shift to the left, too. The shift in marginal revenue will change the profit-maximizing quantity that the firm chooses to produce, since marginal revenue will then equal marginal cost at a lower quantity.

Figure 10.4 (a) shows a situation in which a monopolistic competitor was earning a profit with its original perceived demand curve (D0). The intersection of the marginal revenue curve (MR0) and marginal cost curve (MC) occurs at point S, corresponding to quantity Q0, which is associated on the demand curve at point T with price P0. The combination of price P0 and quantity Q0 lies above the average cost curve, which shows that the firm is earning positive economic profits.

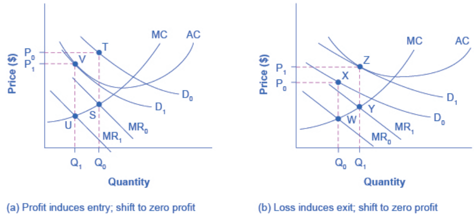


Figure 10.4. *Monopolistic Competition, Entry, and Exit (a) At P0 and Q0, the monopolistically competitive firm shown in this figure is making a positive economic profit. This is clear because if you follow the dotted line above Q0, you can see that price is above average cost. Positive economic profits attract competing firms to the industry, driving the original firm's demand down to D1. At the new equilibrium quantity (P1, Q1), the original firm is earning zero economic profits, and entry into the industry ceases. In (b) the opposite occurs. At P0 and Q0, the firm is losing*

money. If you follow the dotted line above Q0, *you can see that average cost is above price. Losses induce firms to leave the industry. When they do, demand for the original firm rises to D1, where once again the firm is earning zero economic profit.*

Unlike a monopoly, with its high barriers to entry, a monopolistically competitive firm with positive economic profits will attract competition. When another competitor enters the market, the original firm's perceived demand curve shifts to the left, from D0 to D1, and the associated marginal revenue curve shifts from MR0 to MR1 (as shown in figure 10.4a). The new profit-maximizing output is Q1, because the intersection of the MR1 and MC now occurs at point U. Moving vertically up from that quantity on the new demand curve, the optimal price is at P1.

As long as the ***firm*** is earning positive economic profits, new competitors will continue to enter the market, reducing the original firm's demand and marginal revenue curves. The long-run ***equilibrium*** is shown in the figure at point V, where the firm's perceived demand curve touches the average cost curve. When price is equal to average cost, economic profits are zero. Thus, although a monopolistically competitive firm may earn positive economic profits in the short term, the process of new entry will drive down economic profits to zero in the long run. Remember that zero economic profit is not equivalent to zero ***accounting profit***. A zero economic profit means the firm's accounting profit is equal to what its resources could earn in their next best use. Figure 10.4 (b) shows the reverse situation, where a monopolistically competitive firm is originally losing money. The adjustment to long-run equilibrium is analogous to the previous example. The economic losses lead to firms exiting, which will result in increased demand for this particular firm, and consequently lower losses. Firms exit up to the point where there are no more losses in this market, for example when the demand curve touches the average cost curve, as in point Z.

Monopolistic competitors can make an economic profit or loss in the short run, but in the long run, entry and exit will drive these firms toward a zero economic profit outcome. However, the zero economic profit outcome in monopolistic competition looks different from the zero economic profit outcome in perfect competition in several ways relating both to efficiency and to variety in the market.

Self Check: Short Run and Long Run Equilibrium

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/684

216. Outcome: Advertising

What you'll learn to do: understand how product differentiation works in monopolistically competitive industries and how firms use advertising to differentiate their products, understanding impact on elasticity

In this section, you will see how firms strive hard to emphasize the dissimilarities between their products and other products in the industry.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Advertising and Monopolistic Competition
- Self Check: Advertising

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

217. Reading: Advertising and Monopolistic Competition

How does advertising impact monopolistic competition?

The U.S. economy spent about $139.5 billion on advertising in 2012, according to Kantar Media Reports. Roughly one third of this was television advertising, and another third was divided roughly equally between Internet, newspapers, and radio. The remaining third was divided up between direct mail, magazines, telephone directory yellow pages, billboards, and other miscellaneous sources. More than 500,000 workers held jobs in the advertising industry.

Advertising is all about explaining to people, or making people believe, that the products of one firm are differentiated from the products of another firm. In the framework of monopolistic competition, there are two ways to conceive of how advertising works: either advertising causes a firm's perceived demand curve to become more inelastic (that is, it causes the perceived demand curve to become steeper); or advertising causes demand for the firm's product to increase (that is, it causes the firm's perceived demand curve to shift to the right). In either case, a successful advertising campaign may allow a firm to sell either a greater quantity or to charge a higher price, or both, and thus increase its profits.

However, economists and business owners have also long suspected that much of the advertising may only offset other advertising. Economist A. C. Pigou wrote the following back in 1920 in his book, *The Economics of Welfare*:

> It may happen that expenditures on advertisement made

> by competing monopolists [that is, what we now call monopolistic competitors] will simply neutralise one another, and leave the industrial position exactly as it would have been if neither had expended anything. For, clearly, if each of two rivals makes equal efforts to attract the favour of the public away from the other, the total result is the same as it would have been if neither had made any effort at all.

Watch this video about the success Axe bodyspray has seen through its advertising campaign:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=246#oembed-1

Self Check: Advertising

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/685

218. Outcome: Inefficiency of Monopolistic Competition

What you'll learn to do: understand why monopolistically competitive markets are inefficient (including deadweight loss)

In this section, you will see why monopolistically competitive markets are not productively efficient.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Monopolistic Competition and Efficiency
- Self Check: Inefficiency of Monopolistic Competition

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

219. Reading: Monopolistic Competition and Efficiency

Monopolistic Competition and Efficiency

The long-term result of entry and exit in a perfectly competitive market is that all firms end up selling at the price level determined by the lowest point on the average cost curve. This outcome is why perfect competition displays ***productive efficiency***: goods are being produced at the lowest possible average cost. However, in monopolistic competition, the end result of entry and exit is that firms end up with a price that lies on the downward-sloping portion of the average cost curve, not at the very bottom of the AC curve. Thus, monopolistic competition will not be productively efficient.

In a perfectly competitive market, each firm produces at a quantity where price is set equal to marginal cost, both in the short run and in the long run. This outcome is why perfect competition displays allocative efficiency: the social benefits of additional production, as measured by the marginal benefit, which is the same as the price, equal the marginal costs to society of that production. In a monopolistically competitive market, the rule for maximizing profit is to set MR = MC—and price is higher than marginal revenue, not equal to it because the demand curve is downward sloping. When P > MC, which is the outcome in a monopolistically competitive market, the benefits to society of providing additional quantity, as measured by the price that people are willing to pay, exceed the marginal costs to society of producing those units. A monopolistically competitive firm does not produce more, which means that society loses the net benefit of those extra units. This is the same argument we made about monopoly, but in this case to a lesser degree. Thus, a monopolistically competitive industry will produce a lower quantity of a good and charge a higher price for it than would a perfectly competitive industry.

Why Does a Shift in Perceived Demand Cause a Shift in Marginal Revenue?

The combinations of price and quantity at each point on a firm's perceived demand curve are used to calculate total revenue for each combination of price and quantity. This information on total revenue is then used to calculate marginal revenue, which is the change in total revenue divided by the change in quantity. A change in perceived demand will change total revenue at every quantity of output and in turn, the change in total revenue will shift marginal revenue at each quantity of output. Thus, when entry occurs in a monopolistically competitive industry, the perceived demand curve for each firm will shift to the left, because a smaller quantity will be demanded at any given price. Another way of interpreting this shift in demand is to notice that, for each quantity sold, a lower price will be charged. Consequently, the marginal revenue will be lower for each quantity sold—and the marginal revenue curve will shift to the left as well. Conversely, exit causes the perceived demand curve for a monopolistically competitive firm to shift to the right and the corresponding marginal revenue curve to shift right, too.

A monopolistically competitive industry does not display productive and allocative efficiency in either the short run, when firms are making economic profits and losses, nor in the long run, when firms are earning zero profits.

THE BENEFITS OF VARIETY AND PRODUCT DIFFERENTIATION

Even though monopolistic competition does not provide productive efficiency or allocative efficiency, it does have benefits of its own. Product differentiation is based on variety and innovation. Many people would prefer to live in an economy with many kinds of clothes, foods, and car styles; not in a world of perfect competition where everyone will always wear blue jeans and white shirts, eat only spaghetti with plain red sauce, and drive an identical model of car. Many people would prefer to live in an economy where firms are struggling to figure out ways of attracting customers by methods like friendlier service, free delivery, guarantees of quality, variations on existing products, and a better shopping experience.

Economists have struggled, with only partial success, to address the question of whether a market-oriented economy produces the optimal amount of variety. Critics of market-oriented economies argue that society does not really need dozens of different athletic shoes or breakfast cereals or automobiles. They argue that much of the cost of creating such a high degree of product differentiation, and then of advertising and marketing this differentiation, is socially wasteful—that is, most people would be just as happy with a smaller range of ***differentiated products*** produced and sold at a lower price. Defenders of a market-oriented economy respond that if people do not want to buy differentiated products or highly advertised brand names, no one is forcing them to do so. Moreover, they argue that consumers benefit substantially when firms seek short-term profits by providing differentiated products. This controversy may never be fully resolved, in part because deciding on the optimal amount of variety is very difficult, and in part because the two sides often place different values on what variety means for consumers.

Self Check: Inefficiency of Monopolistic

Competition

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/686

220. Putting It Together: Monopolistically Competitive Industries

Summary

The goal of this module was analyze a firm's profit maximizing strategies under conditions of monopolistic competition. You learned how to:

- Define the characteristics of a monopolistically competitive industry
- Calculate and graph the firm's fixed, variable, average, marginal and total costs
- Explain the difference between short run and long run equilibrium in a monopolistically competitive industry
- Understand how product differentiation works in monopolistically competitive industries and how firms use advertising to differentiate their products, understanding impact on elasticity
- Understand why monopolistically competitive markets are inefficient (including deadweight loss)

Examples

Monopolistically competitive industries consist of a significant number of firms, which each produce a differentiated (or heterogeneous) production. In other words, Colgate, AIM, and Tom's of Maine each produce toothpaste, but they are not identical products. Like firms in any market structure, if a monopolistically competitive firm wishes to maximize profits, it will supply the quantity of output where marginal revenue equals marginal cost. Like perfectly competitive firms, competition prevents monopolistically competitive firms from earning positive economic profits in the long run, unless those firms create innovative new products and/or use advertising to convince customers they have done so.

Big Mac versus Whopper. *Big Mac Snip by Ian Burt, CC-BY. Burger King Whopper by Mike Mozart, CC-BY.*

Returning to some of the questions posed in the "Why it Matters" feature at the beginning of this module:

- Why do gas stations charge different prices for a gallon of gasoline? Some gasoline companies use different additives to make their products at least appear different. This allows them to charge higher prices than companies that don't make as good a case for their product.
- What determines how far apart the prices of Colgate and Crest toothpaste can be? Location also matters. A gas station just off the highway can charge higher prices than stations further away, because travelers perceive and are willing to pay for the

convenience of the former.

- Why did fast food restaurants start offering salads? Fast food restaurants added salads to their menus to differentiate their product by appealing to health conscious diners.
- Why did McDonalds come up with the Big Mac sandwich? McDonalds invented the Big Mac because its competitors offered similar enough regular burgers that McDonalds lost its monopoly profits. The Big Mac restored those profits, at least until Burger King came up with the Whopper and other fast food restaurants developed their own special burgers.

221. Glossary: Monopolistic Competition

differentiated product
: a product that is perceived by consumers as distinctive in some way

imperfectly competitive
: firms and organizations that fall between the extremes of monopoly and perfect competition

monopolistic competition
: many firms competing to sell similar but differentiated products

oligopoly
: when a few large firms have all or most of the sales in an industry

222. Discussion: Monopolistic Competition

Suppose you manage a local grocery store, and you learn that a very popular national grocery chain is about to open a store just a few miles away. Use the model of monopolistic competition to analyze the impact of this new store on the quantity of output your store should produce (Q) and the price your store should charge (P). What will happen to your profits? Please show graphically and explain your reasoning in detail. For example, how and why do profits change? How can that be seen on the graph?

What could you do to defend your market share against the new store?

PART XIII

MODULE: OLIGOPOLY

223. Why It Matters: Oligopoly

Why analyze a firm's profit maximizing strategies under conditions of oligopoly?

Think about the purchases you make. Perhaps you're buying groceries. Perhaps you're going out to the movies or dinner. Maybe you're making airline reservations for a trip. Maybe you're buying a new car. Almost certainly, the industries you are doing business with are imperfectly competitive. They consist of more than one firm, but less than the large number required for perfect competition. Each firm has some degree of market power, but none can ignore what other firms in the industry are doing.

Airline Collage. *United by Lars Steffans, CC-BY-SA. American Airlines 2 by Aero Icarus, CC-BY-SA.. Delta by Ikarasra, CC-BY. Southwest Airlines by Aero Icarus, CC-BY-SA.*

Imperfectly competitive industries fall into one of two categories, either oligopoly or monopolistic competition. In this module, we discuss oligopoly, the market structure which is most like monopoly. In the next module, we'll discuss monopolistic competition, which is closer to perfect competition.

Most of the firms that get talked about as "monopolies" today or that regulatory authorities pursue antitrust activities against are actually oligopolies, firms that have only a limited number of competitors. There are quite a few industries in the U.S. that are oligopolistic. Think about rental cars, or car manufacturers, or newspapers, or internet service providers.

Here are some questions to think about as you work through this module:

- Why does an oligopoly like Microsoft make large economic profits, while an oligopoly like United Airlines barely breaks even?
- Why are economy seats essentially the same price as they were a decade or two ago, when the prices of most things have gone up?
- Why do airlines offer frequent flyer clubs?

These are just some of the questions you'll be able to answer by the end of our study of imperfect competition.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=253

LEARNING OUTCOMES

- Define characteristics of oligopolies
- Explain why collusion can occur in oligopolistic industries
- Explain the role of game theory in understanding the behavior of oligopolies
- Explain why oligopolies are inefficient

224. Outcome: Introduction to Oligopolies

What you'll learn to do: define characteristics of oligopolies

In this section, you will learn what oligopolies are and why they exist. You'll read about how some oligopolies are motivated to work together and collude to ensure higher profits, while others compete and act more like perfect competitors.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Introducing Oligopololies
- Reading: Why do Oligopolies Exist?
- Reading: Competition Among the Few
- Reading: Oligopoly Models
- Self Check: Introduction to Oligopolies

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

225. Reading: Introducing Oligopolies

THE TEMPTATION TO DEFY THE LAW

Figure 10.1. *Competing Brands? The laundry detergent market is one that is characterized neither as perfect competition nor monopoly. (Credit: modification of work by Pixel Drip/Flickr Creative Commons)*

Laundry detergent and bags of ice—products of industries that seem pretty mundane, maybe even boring. Hardly! Both have been the center of clandestine meetings and secret deals worthy of a spy novel. In France, between 1997 and 2004, the top four laundry detergent producers (Proctor & Gamble, Henkel, Unilever, and Colgate-Palmolive) controlled about 90 percent of the French soap market. Officials from the soap firms were meeting secretly, in out-of-the-way, small cafés around Paris. Their goals: Stamp out competition and set prices.

Around the same time, the top five Midwest ice makers (Home City Ice, Lang Ice, Tinley Ice, Sisler's Dairy, and Products of Ohio) had similar goals in mind when they secretly agreed to divide up the bagged ice market.

If both groups could meet their goals, it would enable each to act as though they were a single firm—in essence, a monopoly—and enjoy monopoly-size profits. The problem? In many parts of the world, including the European Union and the United States, it is illegal for firms to divide up markets and set prices collaboratively.

These two cases provide examples of markets that are characterized neither as perfect competition nor monopoly. Instead, these firms are competing in market structures that lie between the extremes of monopoly and perfect competition. How do they behave? Why do they exist?

Introduction to Oligopoly

One type of imperfectly competitive market is *oligopoly*. Oligopolistic markets are those dominated by a small number of firms. Commercial aircraft provides a good example: Boeing and Airbus each produce slightly less than 50% of the large commercial aircraft in the world. Another example is the U.S. soft drink industry, which is dominated by Coca-Cola and Pepsi. Oligopolies are characterized by high barriers to entry with firms choosing output, pricing, and other decisions strategically based on the decisions of the other firms in the market. In this module, we will discuss oligopolistic firms, which face two conflicting temptations: to collaborate as if they were a single monopoly, or to individually compete to gain profits by expanding output levels and cutting prices. Oligopolistic markets and firms can also take on elements of monopoly and of perfect competition.

226. Reading: Why do Oligopolies Exist?

WHY DO OLIGOPOLIES EXIST?

Many purchases that individuals make at the retail level are produced in markets that are neither perfectly competitive, monopolies, nor monopolistically competitive. Rather, they are oligopolies. ***Oligopoly*** arises when a small number of large firms have all or most of the sales in an industry. Examples of oligopoly abound and include the auto industry, cable television, and commercial air travel. Oligopolistic firms are like cats in a bag. They can either scratch each other to pieces or cuddle up and get comfortable with one another. If oligopolists compete hard, they may end up acting very much like perfect competitors, driving down costs and leading to zero profits for all. If oligopolists collude with each other, they may effectively act like a monopoly and succeed in pushing up prices and earning consistently high levels of profit. Oligopolies are typically characterized by mutual interdependence where various decisions such as output, price, advertising, and so on, depend on the decisions of the other firm(s). Analyzing the choices of oligopolistic firms about pricing and quantity produced involves considering the pros and cons of competition versus collusion at a given point in time.

A combination of the barriers to entry that create monopolies and the product differentiation that characterizes monopolistic competition can create the setting for an oligopoly. For example, when a government grants a patent for an invention to one firm, it may create a monopoly. When the government grants patents to, for example, three different pharmaceutical companies that each has its own drug for reducing high blood pressure, those three firms may become an oligopoly.

Similarly, a ***natural monopoly*** will arise when the quantity demanded in a market is only large enough for a single firm to operate at the minimum of the long-run average cost curve. In such a setting, the market has room for only one firm, because no smaller firm can operate at a low enough average cost to compete, and no larger firm could sell what it produced given the quantity demanded in the market.

Quantity demanded in the market may also be two or three times the quantity needed to produce at the minimum of the average cost curve—which means that the market would have room for only two or three oligopoly firms (and they need not produce differentiated products). Again, smaller firms would have higher average costs and be unable to compete, while additional large firms would produce such a high quantity that they would not be able to sell it at a profitable price. This combination of economies of scale and market demand creates the barrier to entry, which led to the Boeing-Airbus oligopoly for large passenger aircraft.

The product differentiation at the heart of monopolistic competition can also play a role in creating oligopoly. For example, firms may need to reach a certain minimum size before they are able to spend enough on advertising and marketing to create a recognizable brand name. The problem in competing with, say, Coca-Cola or Pepsi is not that producing fizzy drinks is technologically difficult, but rather that creating a brand name and marketing effort to equal Coke or Pepsi is an enormous task.

227. Reading: Competition Among the Few

Competition Among the Few

In July, 2005, General Motors Corporation (GMC) offered "employee discount pricing" to virtually all GMC customers, not just employees and their relatives. This new marketing strategy introduced by GMC obviously affected Ford, Chrysler, Toyota and other automobile and truck manufacturers; Ford matched GMC's employee-discount plan by offering up to $1,000 to its own employees who convinced friends to purchase its cars and trucks. Ford also offered its customers the same prices paid by its employees. By mid-July, Chrysler indicated that it was looking at many alternatives, but was waiting for GMC to make its next move. Ultimately, Chrysler also offered employee discount pricing.

Toyota had to respond. It quickly developed a new marketing strategy of its own, which included lowering the prices of its cars and offering new financing terms. The responses of Ford, Chrysler, and Toyota to GMC's pricing strategy obviously affected the outcome of that strategy. Similarly, a decision by Procter & Gamble to lower the price of Crest toothpaste may elicit a response from Colgate-Palmolive, and that response will affect the sales of Crest. In an oligopoly, the market is dominated by a few firms, each of which recognizes that its own actions will produce a response from its rivals and that those responses will affect it.

The firms that dominate an oligopoly recognize that they are interdependent: What one firm does affects each of the others. This interdependence stands in sharp contrast to the models of perfect competition and monopolistic competition, where we assume that each firm is so small that it assumes the rest of the market will, in effect, ignore what it does. A perfectly competitive firm responds to the market, not to the actions of any other firm. A monopolistically competitive firm responds to its own demand, not to the actions of specific rivals. These presumptions greatly simplify the analysis of perfect competition and monopolistic competition. We do not have that luxury in oligopoly, where the interdependence of firms is the defining characteristic of the market.

Some oligopoly industries make standardized products: steel, aluminum, wire, and industrial tools. Others make differentiated products: cigarettes, automobiles, computers, ready-to-eat breakfast cereal, and soft drinks.

Measuring Concentration in Oligopoly

Oligopoly means that a few firms dominate an industry. But how many is "a few," and how large a share of industry output does it take to "dominate" the industry?

Compare, for example, the ready-to-eat breakfast cereal industry and the ice cream industry. The cereal market is dominated by two firms, Kellogg's and General Mills, which together hold more than half the cereal market. This oligopoly operates in a highly concentrated market. The market for ice cream, where the four largest firms account for just less than a third of output, is much less concentrated.

One way to measure the degree to which output in an industry is concentrated among a few firms is to use a concentration ratio, which reports the percentage of output accounted for by the largest firms in an industry. The higher the concentration ratio, the more the firms in the industry take account of their rivals' behavior. The lower the concentration ratio, the more the industry reflects the characteristics of monopolistic competition or perfect competition.

The U.S. Census Bureau, based on surveys it conducts of manufacturing firms every five years, reports concentration ratios.

These surveys show concentration ratios for the largest 4, 8, 20, and 50 firms in each industry category. Some concentration ratios from the 2007 survey, the latest available, are reported in Table 11.1 "Concentration Ratios and Herfindahl–Hirschman Indexes." Notice that the four-firm concentration ratio for breakfast cereals is 80%; for ice cream it is 53%.

Table 11.1 Concentration Ratios and Herfindahl–Hirschman Indexes

Industry	**Largest 4 firms**	**Largest 8 firms**	**Largest 20 firms**	**Largest 50 firms**	**HHI**
Ice cream	53	66	84	94	954
Breakfast cereals	80	92	100	100	2426
Cigarettes	98	99	100		*D
Men's and boys' shirts	56	75	90	98	1102
Women's and girls' blouses and shirts	42	58	80	94	719
Automobiles	68	91	99	100	1449
Sporting and athletic goods	27	38	53	68	253
Dental laboratories	18	24	29	36	102
*D, data withheld by the government to avoid revealing information about specific firms.					

Two measures of industry concentration are reported by the Census Bureau: concentration ratios and the Herfindahl–Hirschman Index (HHI). *Source:* *[Selected statistics from Sector 31](#)*: *Manufacturing: Subject Series—Concentration Ratios: Share of Value of Shipments Accounted for by the* 4, 8, 20, *and* 50 *Largest Companies for Industries:* 2007.

An alternative measure of concentration is found by squaring the percentage share (stated as a whole number) of each firm in an industry, then summing these squared market shares to derive a Herfindahl–Hirschman Index (HHI). The largest HHI possible is the case of monopoly, where one firm has 100% of the market; the index is 100^2, or 10,000. An industry with two firms, each with 50% of total output, has an HHI of 5,000 ($50^2 + 50^2$). In an industry with 10,000 firms that have 0.01% of the market each, the HHI is 1. Herfindahl–Hirschman Indexes reported by the Census Bureau are also given in Table 11.1 "Concentration Ratios and Herfindahl–Hirschman Indexes." Notice that the HHI is 2,521 for breakfast cereals and only 736 for ice cream, suggesting that the ice cream industry is more competitive than the breakfast cereal industry.

In some cases, the census data understate the degree to which a few firms dominate the market. One problem is that industry categories may be too broad to capture significant cases of industry dominance. The sporting goods industry, for example, appears to be highly competitive if we look just at measures of concentration, but markets for individual goods, such as golf clubs, running shoes, and tennis rackets, tend to be dominated by a few firms. Further, the data reflect shares of the national market. A tendency for regional domination does not show up. For example, the concrete industry appears to be highly competitive. But concrete is produced in local markets—it is too expensive to ship it very far—and many of these local markets are dominated by a handful of firms.

The census data can also overstate the degree of actual concentration. The "automobiles" category, for example, has a four-firm concentration ratio that suggests the industry is strongly dominated by four large firms (in fact, U.S. production is dominated by three: General Motors, Ford, and Chrysler). Those firms hardly account for all car sales in the United States, however, as other foreign producers have captured a large portion of the domestic market. Including those foreign competitors suggests a far less concentrated industry than the census data imply.

228. Reading: Oligopoly Models

Oligopoly Models

There is no generally accepted model of oligopoly, but rather there are a number of models that will be touched on in the following sections. In principle, one can calculate and graph an oligopoly's cost and revenue curves, and determine its profit maximizing level of output and price in the same way as we did with monopoly. What complicates matters with oligopolistic industries is that any one firm's demand and marginal revenue curves are influenced by what the other oligopolistic firms are doing. For example, if Pepsi goes on sale, the demand for Coca-Cola declines, with the demand and marginal revenue curves shifting to the left. Thus, the best price and quantity for any oligopoly depends on what every other firm in the industry is doing.

Self Check: Introduction to Oligopolies

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the four Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/673

229. Outcome: Collusion

What you'll learn to do: explain why collusion can occur in oligopolistic industries

In this learning outcome, you will learn about how a firm in an oligopoly is enticed to collaborate with other firms so that they can collectively control the market.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Collusion or Competition?
- Reading: The Collusion Model
- Self Check: Collusion

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

230. Reading: Collusion or Competition?

Collusion versus cartels: How can I tell which is which?

When oligopoly firms in a certain market decide what quantity to produce and what price to charge, they face a temptation to act as if they were a monopoly. By acting together, oligopolistic firms can hold down industry output, charge a higher price, and divide up the profit among themselves. When firms act together in this way to reduce output and keep prices high, it is called ***collusion***. A group of firms that have a formal agreement to collude to produce the monopoly output and sell at the monopoly price is called a ***cartel***.

In the United States, as well as many other countries, it is illegal for firms to collude since collusion is anti-competitive behavior, which is a violation of antitrust law. Both the *Antitrust Division of the Justice Department* and the *Federal Trade Commission* have responsibilities for preventing collusion in the United States.

The problem of enforcement is finding hard evidence of collusion. Cartels are formal agreements to collude. Because cartel agreements provide evidence of collusion, they are rare in the United States. Instead, most collusion is tacit, where firms implicitly reach an understanding that competition is bad for profits.

The desire of businesses to avoid competing so that they can instead raise the prices that they charge and earn higher profits has been well understood by economists. *Adam Smith* wrote in *Wealth of Nations* in 1776: "People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices."

Even when oligopolists recognize that they would benefit as a group by acting like a monopoly, each individual oligopoly faces a private temptation to produce just a slightly higher quantity and earn slightly higher profit—while still counting on the other oligopolists to hold down their production and keep prices high. If at least some oligopolists give in to this temptation and start producing more, then the market price will fall. Indeed, a small handful of oligopoly firms may end up competing so fiercely that they all end up earning zero economic profits—as if they were perfect competitors.

231. Reading: The Collusion Model

The Collusion Model

There is no single model of profit-maximizing oligopoly behavior that corresponds to economists' models of perfect competition, monopoly, and monopolistic competition. Uncertainty about the interaction of rival firms makes specification of a single model of oligopoly impossible. Instead, economists have devised a variety of models that deal with the uncertain nature of rivals' responses in different ways. In this section we review one type of oligopoly model, the collusion model. After examining this traditional approach to the analysis of oligopoly behavior, we shall turn to another method of examining oligopolistic interaction: game theory.

Firms in any industry could achieve the maximum profit attainable if they all agreed to select the monopoly price and output and to share the profits. One approach to the analysis of oligopoly is to assume that firms in the industry collude, selecting the monopoly solution.

Suppose an industry is a duopoly, an industry with two firms. Figure 11.3 "Monopoly Through Collusion" shows a case in which the two firms are identical. They sell identical products and face identical demand and cost conditions. To simplify the analysis, we will assume that each has a horizontal marginal cost curve, *MC*. The demand and marginal revenue curves are the same for both firms. We find the combined demand curve for the two firms, D_{combined}, by adding the individual demand curves together. Because one firm's demand curve, D_{firm}, represents one-half of market demand, it is the same as the combined marginal revenue curve for the two firms. If these two firms act as a monopoly, together they produce

Q_m and charge a price P_m. This result is achieved if each firm selects its profit-maximizing output, which equals 1/2 Q_m. This solution is inefficient; the efficient solution is price P_c and output Q_c, found where the combined market demand curve$D_{combined}$ and the marginal cost curve MC intersect.

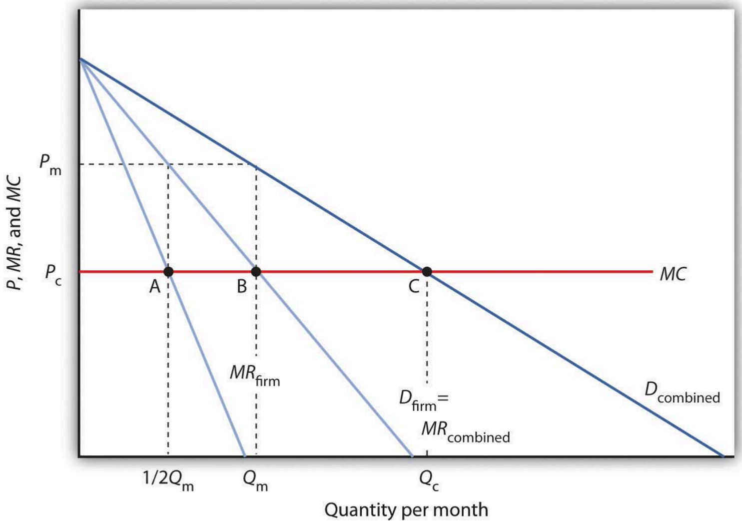


Figure 11.3. *Monopoly Through Collusion. Two identical firms have the same horizontal marginal cost curve MC. Their demand curves Dfirm and marginal revenue curves MRfirm are also identical. The combined demand curve is Dcombined; the combined marginal revenue curve is MRcombined. The profits of the two firms are maximized if each produces 1/2 Qm at point A. Industry output at point B is thus Qm and the price is Pm. At point C, the efficient solution output would be Qc, and the price would equal MC.*

In the simplest form of collusion, overt collusion, firms openly agree on price, output, and other decisions aimed at achieving monopoly profits. Firms that coordinate their activities through overt collusion and by forming collusive coordinating mechanisms make up a cartel.

Firms form a cartel to gain monopoly power. A successful cartel can earn large profits, but there are several problems with forming and maintaining one. First, in many countries, including the United States, cartels are generally illegal.[1] They are banned, because their purpose is to raise prices and restrict output. Second, the cartel may not succeed in inducing all firms in the industry to join. Firms that remain outside the cartel can compete by lowering price, and thus they prevent the cartel from achieving the monopoly solution. Third, there is always an incentive for individual members to cheat on cartel agreements. Suppose the members of a cartel have agreed to impose the monopoly price in their market and to limit their output accordingly. Any one firm might calculate that it could charge slightly less than the cartel price and thus capture a larger share of the market for itself. Cheating firms expand output and drive prices down below the level originally chosen.

The Organization of Petroleum Exporting Countries (OPEC), perhaps the best-known cartel, is made up of 13 oil-producing countries. In the 1970s, OPEC successfully acted like a monopoly by restricting output and raising prices. By the mid-1980s, however, the monopoly power of the cartel had been weakened by expansion of output by nonmember producers such as Mexico and Norway and by cheating among the cartel members.

An alternative to overt collusion is tacit collusion, an unwritten, unspoken understanding through which firms agree to limit their competition. Firms may, for example, begin following the price leadership of a particular firm, raising or lowering their prices when the leader makes such a change. The price leader may be the largest firm in the industry, or it may be a firm that has been particularly good at assessing changes in demand or cost. At various times, tacit

1. One legal cartel is the NCAA, which many economists regard as a successful device through which member firms (colleges and universities) collude on a wide range of rules through which they produce sports.

collusion has been alleged to occur in a wide range of industries, including steel, cars, and breakfast cereals.

It is difficult to know how common tacit collusion is. The fact that one firm changes its price shortly after another one does cannot prove that a tacit conspiracy exists. After all, we expect to see the prices of all firms in a perfectly competitive industry moving together in response to changes in demand or production costs.

Self Check: Collusion

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/674

232. Outcome: Game Theory

What you'll learn to do: explain the role of game theory in understanding the behavior of oligopolies

In this outcome, you will analyze the fascinating dilemmas firms may face in an oligopoly—cooperate or act independently?

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Prisoner's Dilemma
- Reading: Game Theory
- Self Check: Game Theory

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

233. Reading: Prisoner's Dilemma

Prisoner's Dilemma

Because of the complexity of oligopoly, which is the result of mutual interdependence among firms, there is no single, generally-accepted theory of how oligopolies behave, in the same way that we have theories for all the other market structures. Instead, economists use ***game theory***, a branch of mathematics that analyzes situations in which players must make decisions and then receive payoffs based on what other players decide to do. Game theory has found widespread applications in the social sciences, as well as in business, law, and military strategy.

The ***prisoner's dilemma*** is a scenario in which the gains from cooperation are larger than the rewards from pursuing self-interest. It applies well to oligopoly. The story behind the prisoner's dilemma goes like this:

> Two co-conspiratorial criminals are arrested. When they are taken to the police station, they refuse to say anything and are put in separate interrogation rooms. Eventually, a police officer enters the room where Prisoner A is being held and says: "You know what? Your partner in the other room is confessing. So your partner is going to get a light prison sentence of just one year, and because you're remaining silent, the judge is going to stick you with eight years in prison. Why don't you get smart? If you confess, too, we'll cut your jail time down to five years, and your partner will get five years, also." Over in the next room, another police officer is giving exactly the same speech to Prisoner B. What

> the police officers do not say is that if both prisoners remain silent, the evidence against them is not especially strong, and the prisoners will end up with only two years in jail each.

The game theory situation facing the two prisoners is shown in Table 10.2. To understand the dilemma, first consider the choices from Prisoner A's point of view. If A believes that B will confess, then A ought to confess, too, so as to not get stuck with the eight years in prison. But if A believes that B will not confess, then A will be tempted to act selfishly and confess, so as to serve only one year. The key point is that A has an incentive to confess regardless of what choice B makes! B faces the same set of choices, and thus will have an incentive to confess regardless of what choice A makes. Confess is considered the dominant strategy or the strategy an individual (or firm) will pursue regardless of the other individual's (or firm's) decision. The result is that if prisoners pursue their own self-interest, both are likely to confess, and end up doing a total of 10 years of jail time between them.

Table 10.2. The Prisoner's Dilemma Problem

		Prisoner B	
		Remain Silent (cooperate with other prisoner)	Confess (do not cooperate with other prisoner)
Prisoner A	Remain Silent (cooperate with other prisoner)	A gets 2 years, B gets 2 years	A gets 8 years, B gets 1 year
	Confess (do not cooperate with other prisoner)	A gets 1 year, B gets 8 years	A gets 5 years B gets 5 years

The game is called a dilemma because if the two prisoners had cooperated by both remaining silent, they would only have had to serve a total of four years of jail time between them. If the two prisoners can work out some way of cooperating so that neither one will confess, they will both be better off than if they each follow their own individual self-interest, which in this case leads straight into longer jail terms.

THE OLIGOPOLY VERSION OF THE PRISONER'S DILEMMA

The members of an oligopoly can face a prisoner's dilemma, also. If each of the oligopolists cooperates in holding down output, then high monopoly profits are possible. Each oligopolist, however, must worry that while it is holding down output, other firms are taking advantage of the high price by raising output and earning higher profits. Table 10.3 shows the prisoner's dilemma for a two-firm oligopoly—known as a *duopoly*. If Firms A and B both agree to hold down output, they are acting together as a monopoly and will each earn $1,000 in profits. However, both firms' dominant strategy is to increase output, in which case each will earn $400 in profits.

Table 10.3. A Prisoner's Dilemma for Oligopolists

		Firm B	
		Hold Down Output (cooperate with other firm)	Increase Output (do not cooperate with other firm)
Firm A	Hold Down Output (cooperate with other firm)	A gets $1,000, B gets $1,000	A gets $200, B gets $1,500
	Increase Output (do not cooperate with other firm)	A gets $1,500, B gets $200	A gets $400, B gets $400

Can the two firms trust each other? Consider the situation of Firm A:

- If A thinks that B will cheat on their agreement and increase output, then A will increase output, too, because for A the profit of $400 when both firms increase output (the bottom right-hand choice in Table 10.3) is better than a profit of only $200 if A keeps output low and B raises output (the upper right-hand choice in the table).

- If A thinks that B will cooperate by holding down output, then A may seize the opportunity to earn higher profits by raising output. After all, if B is going to hold down output, then A can earn $1,500 in profits by expanding output (the bottom left-hand choice in the table) compared with only $1,000 by holding down output as well (the upper left-hand choice in the table).

Thus, firm A will reason that it makes sense to expand output if B holds down output and that it also makes sense to expand output if B raises output. Again, B faces a parallel set of decisions.

The result of this prisoner's dilemma is often that even though A and B could make the highest combined profits by cooperating in producing a lower level of output and acting like a monopolist, the two firms may well end up in a situation where they each increase *output* and earn only $400 each in *profits*. The following example discusses one cartel scandal in particular.

What is the Lysine cartel?

Lysine, a $600 million-a-year industry, is an amino acid used by farmers as a feed additive to ensure the proper growth of swine and poultry. The primary U.S. producer of lysine is Archer Daniels Midland (ADM), but several other large European and Japanese firms are also in this market. For a time in the first half of the 1990s, the world's major lysine producers met together in hotel conference rooms and decided exactly how much each firm would sell and what it would charge. The U.S. Federal Bureau of Investigation (FBI), however, had learned of the cartel and placed wire taps on a number of their phone calls and meetings.

From FBI surveillance tapes, following is a comment that Terry

Wilson, president of the corn processing division at ADM, made to the other lysine producers at a 1994 meeting in Mona, Hawaii:

> I wanna go back and I wanna say something very simple. If we're going to trust each other, okay, and if I'm assured that I'm gonna get 67,000 tons by the year's end, we're gonna sell it at the prices we agreed to . . . The only thing we need to talk about there because we are gonna get manipulated by these [expletive] buyers—they can be smarter than us if we let them be smarter. . . . They [the customers] are not your friend. They are not my friend. And we gotta have 'em, but they are not my friends. You are my friend. I wanna be closer to you than I am to any customer. Cause you can make us … money. … And all I wanna tell you again is let's—let's put the prices on the board. Let's all agree that's what we're gonna do and then walk out of here and do it.

The price of lysine doubled while the cartel was in effect. Confronted by the FBI tapes, Archer Daniels Midland pled guilty in 1996 and paid a fine of $100 million. A number of top executives, both at ADM and other firms, later paid fines of up to $350,000 and were sentenced to 24–30 months in prison.

In another one of the FBI recordings, the president of Archer Daniels Midland told an executive from another competing firm that ADM had a slogan that, in his words, had "penetrated the whole company." The company president stated the slogan this way: "Our competitors are our friends. Our customers are the enemy." That slogan could stand as the motto of cartels everywhere.

HOW TO ENFORCE COOPERATION

How can parties who find themselves in a prisoner's dilemma situation avoid the undesired outcome and cooperate with each other? The way out of a prisoner's dilemma is to find a way to penalize those who do not cooperate.

Perhaps the easiest approach for colluding oligopolists, as you might imagine, would be to sign a contract with each other that they will hold output low and keep prices high. If a group of U.S. companies signed such a contract, however, it would be illegal. Certain international organizations, like the nations that are members of the *Organization of Petroleum Exporting Countries* (OPEC), have signed international agreements to act like a monopoly, hold down output, and keep prices high so that all of the countries can make high profits from oil exports. Such agreements, however, because they fall in a gray area of international law, are not legally enforceable. If Nigeria, for example, decides to start cutting prices and selling more oil, Saudi Arabia cannot sue Nigeria in court and force it to stop.

LINK IT UP

Visit the Organization of the Petroleum Exporting Countries website and learn more about its history and how it defines itself.

Because oligopolists cannot sign a legally enforceable contract to act like a monopoly, the firms may instead keep close tabs on what other firms are producing and charging. Alternatively, oligopolists may choose to act in a way that generates pressure on each firm to stick to its agreed quantity of output.

One example of the pressure these firms can exert on one another is the ***kinked demand curve***, in which competing oligopoly firms commit to match price cuts, but not price increases. This situation is shown in Figure 10.5. Say that an oligopoly airline has agreed with the rest of a cartel to provide a quantity of 10,000 seats on the New York to Los Angeles route, at a price of $500. This choice defines the kink in the firm's perceived demand curve. The reason that the firm faces a kink in its demand curve is because of how the other oligopolists react to changes in the firm's price. If the oligopoly decides to produce more and cut its price, the other members of the cartel will immediately match any price cuts—and therefore, a lower price brings very little increase in quantity sold.

If one *firm* cuts its price to $300, it will be able to sell only 11,000 seats. However, if the airline seeks to raise prices, the other oligopolists will not raise their prices, and so the firm that raised prices will lose a considerable share of sales. For example, if the firm raises its price to $550, its sales drop to 5,000 seats sold. Thus, if oligopolists always match price cuts by other firms in the cartel, but do not match price increases, then none of the oligopolists will have a strong incentive to change prices, since the potential gains are minimal. This strategy can work like a silent form of cooperation, in which the cartel successfully manages to hold down output, increase *price*, and share a monopoly level of profits even without any legally enforceable agreement.

Figure 10.5. A Kinked Demand Curve. Consider a member firm in an oligopoly cartel that is supposed to produce a quantity of 10,000 and sell at a price of $500. The other members of the cartel can encourage this firm to honor its commitments by acting so that the firm faces a kinked demand curve. If the oligopolist attempts to expand output and reduce price slightly, other firms also cut prices immediately—so if the firm expands output to 11,000, the price per unit falls dramatically, to $300. On the other side, if the oligopoly attempts to raise its price, other firms will not do so, so if the firm raises its price to $550, its sales decline sharply to 5,000. Thus, the members of a cartel can discipline each other to stick to the pre-agreed levels of quantity and price through a strategy of matching all price cuts but not matching any price increases.

Many real-world oligopolies, prodded by economic changes, legal and political pressures, and the egos of their top executives, go through episodes of cooperation and competition. If oligopolies could sustain cooperation with each other on output and pricing, they could earn profits as if they were a single monopoly. However, each firm in an oligopoly has an incentive to produce more and grab a bigger share of the overall market; when firms start behaving in this way, the market outcome in terms of prices and quantity can be similar to that of a highly competitive market.

234. Reading: Game Theory

Game Theory and Oligopoly Behavior

Oligopoly presents a problem in which decision makers must select strategies by taking into account the responses of their rivals, which they cannot know for sure in advance. The Start Up feature at the beginning of this module suggested the uncertainty eBay faces as it considers the possibility of competition from Google. A choice based on the recognition that the actions of others will affect the outcome of the choice and that takes these possible actions into account is called a strategic choice. Game theory is an analytical approach through which strategic choices can be assessed.

Among the strategic choices available to an oligopoly firm are pricing choices, marketing strategies, and product-development efforts. An airline's decision to raise or lower its fares—or to leave them unchanged—is a strategic choice. The other airlines' decision to match or ignore their rival's price decision is also a strategic choice. IBM boosted its share in the highly competitive personal computer market in large part because a strategic product-development strategy accelerated the firm's introduction of new products.

Once a firm implements a strategic decision, there will be an outcome. The outcome of a strategic decision is called a payoff. In general, the payoff in an oligopoly game is the change in economic profit to each firm. The firm's payoff depends partly on the strategic choice it makes and partly on the strategic choices of its rivals. Some firms in the airline industry, for example, raised their fares in 2005, expecting to enjoy increased profits as a result. They changed their strategic choices when other airlines chose to slash their fares, and all firms ended up with a payoff of lower profits—many went into bankruptcy.

We shall use two applications to examine the basic concepts of game theory. The first examines a classic game theory problem called the prisoners' dilemma. The second deals with strategic choices by two firms in a duopoly.

The Prisoners' Dilemma

Suppose a local district attorney (DA) is certain that two individuals, Frankie and Johnny, have committed a burglary, but she has no evidence that would be admissible in court.

The DA arrests the two. On being searched, each is discovered to have a small amount of cocaine. The DA now has a sure conviction on a possession of cocaine charge, but she will get a conviction on the burglary charge only if at least one of the prisoners confesses and implicates the other.

The DA decides on a strategy designed to elicit confessions. She separates the two prisoners and then offers each the following deal: "If you confess and your partner doesn't, you will get the minimum sentence of one year in jail on the possession and burglary charges. If you both confess, your sentence will be three years in jail. If your partner confesses and you do not, the plea bargain is off and you will get six years in prison. If neither of you confesses, you will each get two years in prison on the drug charge."

The two prisoners each face a dilemma; they can choose to confess or not confess. Because the prisoners are separated, they cannot plot a joint strategy. Each must make a strategic choice in isolation.

The outcomes of these strategic choices, as outlined by the DA, depend on the strategic choice made by the other prisoner. The payoff matrix for this game is given in Figure 11.6 "Payoff Matrix for the Prisoners' Dilemma". The two rows represent Frankie's strategic choices; she may confess or not confess. The two columns represent Johnny's strategic choices; he may confess or not confess. There are four possible outcomes: Frankie and Johnny both confess (cell A), Frankie confesses but Johnny does not (cell B), Frankie does not confess but Johnny does (cell C), and neither Frankie nor Johnny confesses (cell D). The portion at the lower left in each cell shows Frankie's payoff; the shaded portion at the upper right shows Johnny's payoff.

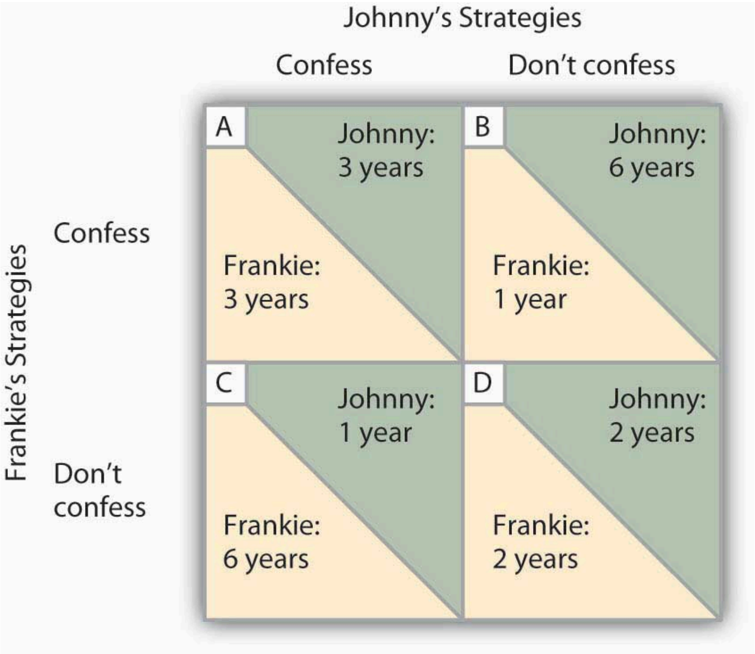


Figure 11.6 *Payoff Matrix for the Prisoners' Dilemma. The four cells represent each of the possible outcomes of the prisoners' game.*

If Johnny confesses, Frankie's best choice is to confess—she will get a three-year sentence rather than the six-year sentence she would get if she did not confess. If Johnny does not confess, Frankie's best strategy is still to confess—she will get a one-year rather than a two-year sentence. In this game, Frankie's best strategy is to confess, regardless of what Johnny does. When a player's best strategy is the same regardless of the action of the other player, that strategy is said to be a dominant strategy. Frankie's dominant strategy is to confess to the burglary.

For Johnny, the best strategy to follow, if Frankie confesses, is to confess. The best strategy to follow if Frankie does not confess is also to confess. Confessing is a dominant strategy for Johnny as well. A game in which there is a dominant strategy for each player is called a dominant strategy equilibrium. Here, the dominant strategy equilibrium is for both prisoners to confess; the payoff will be given by cell A in the payoff matrix.

From the point of view of the two prisoners together, a payoff in cell D would have been preferable. Had they both denied participation in the robbery, their combined sentence would have been four years in prison—two years each. Indeed, cell D offers the lowest combined prison time of any of the outcomes in the payoff matrix. But because the prisoners cannot communicate, each is likely to make a strategic choice that results in a more costly outcome. Of course, the outcome of the game depends on the way the payoff matrix is structured.

Repeated Oligopoly Games

The prisoners' dilemma was played once, by two players. The players were given a payoff matrix; each could make one choice, and the game ended after the first round of choices.

The real world of oligopoly has as many players as there are firms in the industry. They play round after round: a firm raises its price, another firm introduces a new product, the first firm cuts its price, a third firm introduces a new marketing strategy, and so on. An oligopoly game is a bit like a baseball game with an unlimited number of innings—one firm may come out ahead after one round, but another will emerge on top another day. In the computer industry game, the introduction of personal computers changed the rules. IBM, which had won the mainframe game quite handily, struggles to keep up in a world in which rivals continue to slash prices and improve quality.

Oligopoly games may have more than two players, so the games are more complex, but this does not change their basic structure. The fact that the games are repeated introduces new strategic considerations. A player must consider not just the ways in which its choices will affect its rivals now, but how its choices will affect them in the future as well.

We will keep the game simple, however, and consider a duopoly game. The two firms have colluded, either tacitly or overtly, to create a monopoly solution. As long as each player upholds the agreement, the two firms will earn the maximum economic profit possible in the enterprise.

There will, however, be a powerful incentive for each firm to cheat. The monopoly solution may generate the maximum economic profit possible for the two firms combined, but what if one firm captures some of the other firm's profit? Suppose, for example, that two equipment rental firms, Quick Rent and Speedy Rent, operate in a community. Given the economies of scale in the business and the size of the community, it is not likely that another firm will enter. Each firm has about half the market, and they have agreed to charge the prices that would be chosen if the two combined as a single firm. Each earns economic profits of $20,000 per month.

Quick and Speedy could cheat on their arrangement in several ways. One of the firms could slash prices, introduce a new line of rental products, or launch an advertising blitz. This approach would not be likely to increase the total profitability of the two firms, but if one firm could take the other by surprise, it might profit at the expense of its rival, at least for a while.

We will focus on the strategy of cutting prices, which we will call a strategy of cheating on the duopoly agreement. The alternative is not to cheat on the agreement. Cheating increases a firm's profits *if* its rival does not respond. Figure 11.7 "To Cheat or Not to Cheat: Game Theory in Oligopoly" shows the payoff matrix facing the two firms at a particular time. As in the prisoners' dilemma matrix, the four cells list the payoffs for the two firms. If neither firm cheats (cell D), profits remain unchanged.

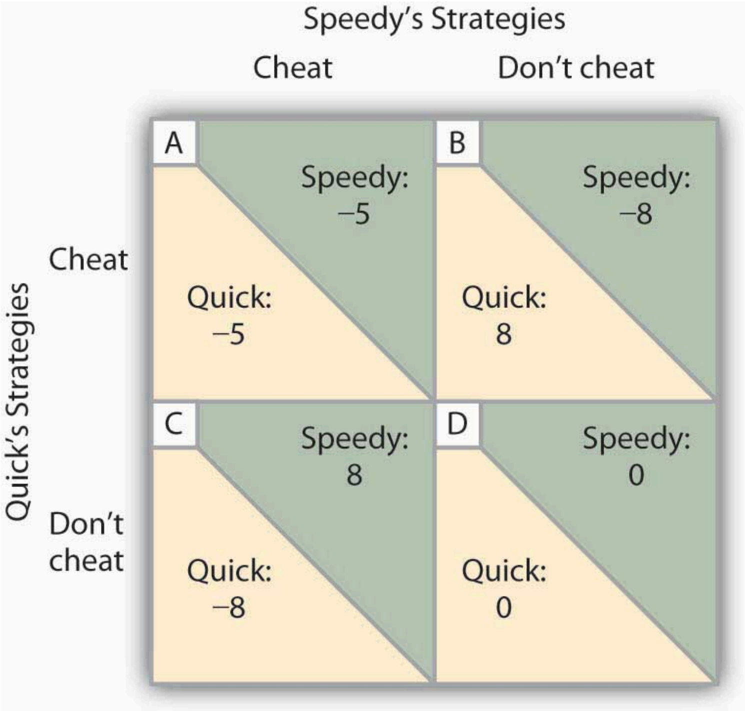


Figure 11.7 To Cheat or Not to Cheat: Game Theory in Oligopoly.

Two rental firms, Quick Rent and Speedy Rent, operate in a duopoly market. They have colluded in the past, achieving a monopoly solution. Cutting prices means cheating on the arrangement; not cheating means maintaining current prices. The payoffs are changes in monthly profits, in thousands of dollars. If neither firm cheats, then neither firm's profits will change. In this game, cheating is a dominant strategy equilibrium.

This game has a dominant strategy equilibrium. Quick's preferred strategy, regardless of what Speedy does, is to cheat. Speedy's best strategy, regardless of what Quick does, is to cheat. The result is that the two firms will select a strategy that lowers their combined profits!

Quick Rent and Speedy Rent face an unpleasant dilemma. They want to maximize profit, yet each is likely to choose a strategy inconsistent with that goal. If they continue the game as it now exists, each will continue to cut prices, eventually driving prices down to the point where price equals average total cost (presumably, the price-cutting will stop there). But that would leave the two firms with zero economic profits.

Both firms have an interest in maintaining the status quo of their collusive agreement. Overt collusion is one device through which the monopoly outcome may be maintained, but that is illegal. One way for the firms to encourage each other not to cheat is to use a tit-for-tat strategy. In a tit-for-tat strategy a firm responds to cheating by cheating, and it responds to cooperative behavior by cooperating. As each firm learns that its rival will respond to cheating by cheating, and to cooperation by cooperating, cheating on agreements becomes less and less likely.

Still another way firms may seek to force rivals to behave cooperatively rather than competitively is to use a trigger strategy, in which a firm makes clear that it is willing and able to respond to cheating by permanently revoking an agreement. A firm might, for example, make a credible threat to cut prices down to the level of average total cost—and leave them there—in response to any price-cutting by a rival. A trigger strategy is calculated to impose huge costs on any firm that cheats—and on the firm that threatens to invoke the trigger. A firm might threaten to invoke a trigger in hopes that the threat will forestall any cheating by its rivals.

Game theory has proved to be an enormously fruitful approach to the analysis of a wide range of problems. Corporations use it to map out strategies and to anticipate rivals' responses. Governments use it in developing foreign-policy strategies. Military leaders play war games on computers using the basic ideas of game theory.

Any situation in which rivals make strategic choices to which competitors will respond can be assessed using game theory analysis.

One rather chilly application of game theory analysis can be found in the period of the Cold War when the United States and the former Soviet Union maintained a nuclear weapons policy that was described by the acronym MAD, which stood for *m*utually *a*ssured *d*estruction. Both countries had enough nuclear weapons to destroy the other several times over, and each threatened to launch sufficient nuclear weapons to destroy the other country if the other country launched a nuclear attack against it or any of its allies. On its face, the MAD doctrine seems, well, mad. It was, after all, a commitment by each nation to respond to any nuclear attack with a counterattack that many scientists expected would end human life on earth. As crazy as it seemed, however, it worked. For 40 years, the two nations did not go to war. While the collapse of the Soviet Union in 1991 ended the need for a MAD doctrine, during the time that the two countries were rivals, MAD was a very effective trigger indeed.

Of course, the ending of the Cold War has not produced the ending of a nuclear threat. Several nations now have nuclear weapons. The threat that Iran will introduce nuclear weapons, given its stated commitment to destroy the state of Israel, suggests that the possibility of nuclear war still haunts the world community.

Self Check: Game Theory

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/675

235. Outcome: Inefficiency in Oligopolies

What you'll learn to do: explain why oligopolies are inefficient

In this section, you will come to see why oligopolies do not efficiently use all of the resources in the market. In an oligopoly, there is typically and underallocation of resources, making oligopolies both productively and allocatively inefficient.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Tradeoffs of Imperfect Competition
- Self Check: Inefficiency in Oligopolies

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

236. Reading: Tradeoffs of Imperfect Competition

Tradeoffs of Imperfect Competition

Oligopoly is probably the second most common market structure (monopolistic competition being the first). When oligopolies result from patented innovations or from taking advantage of economies of scale to produce at low average cost, they may provide considerable benefit to consumers. Oligopolies are often buffeted by significant barriers to entry, which enable the oligopolists to earn sustained profits over long periods of time. Oligopolists also do not typically produce at the minimum of their average cost curves. When they lack vibrant competition, they may lack incentives to provide innovative products and high-quality service.

The task of public policy with regard to competition is to sort through these multiple realities, attempting to encourage behavior that is beneficial to the broader society and to discourage behavior that only adds to the profits of a few large companies, with no corresponding benefit to consumers. Monopoly and Antitrust Policy discusses the delicate judgments that go into this task.

THE TEMPTATION TO DEFY THE LAW

Oligopolistic firms have been called "cats in a bag," as this module mentioned. The French detergent makers chose to "cozy up" with each other. The result? An uneasy and tenuous relationship. When the *Wall Street Journal* reported on the matter, it wrote: "According to a statement a Henkel manager made to the [French anti-trust] commission, the detergent makers wanted 'to limit the intensity of the competition between them and clean up the market.' Nevertheless, by the early 1990s, a price war had broken out among them." During the soap executives' meetings, which sometimes lasted more than four hours, complex pricing structures were established. "One [soap] executive recalled 'chaotic' meetings as each side tried to work out how the other had bent the rules." Like many cartels, the soap cartel disintegrated due to the very strong temptation for each member to maximize its own individual profits.

How did this soap opera end? After an investigation, French antitrust authorities fined Colgate-Palmolive, Henkel, and Proctor & Gamble a total of €361 million ($484 million). A similar fate befell the icemakers. Bagged ice is a commodity, a perfect substitute, generally sold in 7- or 22-pound bags. No one cares what label is on the bag. By agreeing to carve up the ice market, control broad geographic swaths of territory, and set prices, the icemakers moved from perfect competition to a monopoly model. After the agreements, each firm was the sole supplier of bagged ice to a region; there were profits in both the long run and the short run. According to the courts: "These companies illegally conspired to manipulate the marketplace." Fines totaled about $600,000—a steep fine considering a bag of ice sells for under $3 in most parts of the United States.

Even though it is illegal in many parts of the world for firms to set prices and carve up a market, the temptation to earn higher profits makes it extremely tempting to defy the law.

Self Check: Inefficiency in Oligopolies

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/676

237. Putting It Together: Oligopoly

Summary

The goal of this module was analyze a firm's profit maximizing strategies under conditions of oligopoly. You learned how to:

- Define characteristics of oligopolies
- Explain why collusion can occur in oligopolistic industries
- Explain the role of game theory in understanding the behavior of oligopolies
- Explain why oligopolies are inefficient.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=267#oembed-1

Examples

While oligopoly is defined as an industry consisting of, or dominated by a small number of firms, the key characteristic is interdependence among firms. Oligopolies can be characterized by collusion, where firms act jointly like a monopolist to share industry profits, or by competition, where firms compete aggressively for individual profits, or something in between. The computer operating system, dominated by Microsoft, fits the former profile with persistent high economic profits. The airline industry (e.g. United) fits the latter profile, leading to prices barely above costs and low profits.

Oligopolies are inefficient for the same reasons that monopolies are—in order to reap economic profits, they produce too little output so they create deadweight losses to society. The more like a monopoly a given oligopoly is, the higher their profits and the greater the deadweight loss. This is why strong oligopolies usually generate antitrust action by the government.

238. Glossary: Oligopolies

cartel
a group of firms that collude to produce the monopoly output and sell at the monopoly price

collusion
when firms act together to reduce output and keep prices high

duopoly
an oligopoly with only two firms

game theory
a branch of mathematics often used by economists that analyzes situations in which players must make decisions and then receive payoffs based on what decisions the other players make

imperfectly competitive
firms and organizations that fall between the extremes of monopoly and perfect competition

kinked demand curve
a perceived demand curve that arises when competing oligopoly firms commit to match price cuts, but not price increases

oligopoly
when a few large firms have all or most of the sales in an industry

prisoner's dilemma
a game in which the gains from cooperation are larger than the rewards from pursuing self-interest

239. Discussion: Oligopoly

Name a product that you regularly purchase from a firm that operates in an oligopolistic industry. Explain why the product and firm fit the model of oligopoly. Think about the TV commercials and/or print advertisements that you've seen from this industry: What interdependence have you noticed between the firm you selected and its rivals in terms of product differentiation, price leadership, or price competition? Explain your answer.

PART XIV
MODULE: PUBLIC GOODS

240. Why It Matters: Public Goods

Why compare public goods and private goods and understand the role for them in the economy?

Decisions made by firms and individuals in a market often have a spillover effect on other people, whether it be for good (like your neighbor's sweet-smelling cookies), or for the bad (the smelly paper plant that opened up a mile away). In this module, we learn about these positive and negative externalities, as well as the range of public versus private goods.

One of the characteristics of any good or service is its "public" or "private"-ness. Goods can be public or private or anything in between. What does this mean? It's probably not what you think. It doesn't primarily mean who provides it. In other words, a good is not private because it is provided by private businesses. A good is not public because it is provided by the government. Rather, private/public is more a description of how these goods are consumed.

F-15E Strike Eagle Fighter by U.S. Air Force, CC-BY-NC.

All of us consume private goods and public goods. There are three basic cases: A **private good** is one for which the consumer pays all the costs and receives all the benefits. If you buy an apple to eat, no one gets to share any part of the apple that you eat. A **public good** is one where one person's consumption doesn't prevent anyone else from consuming it too.

National defense is the classic example Once it's there, everyone gets the benefits of it. This leads to "free riding," where people try to avoid paying for the public good since they'll get the benefits anyway. In between public and private goods are externality goods (or semi-public goods). Externality goods are characterized by spillovers. People may benefit without paying for externality goods. Or they may have to pay some of the costs without gaining any of the benefits. One example is public K-12 education. Children benefit without paying. Even their parents don't pay the full cost of their education, which tends to be paid for with property taxes regardless of whether the property owner has children in school.

Vintage Class Photo, Globe Public School by freeparking, CC-BY-NC.

- What are some other examples of private goods? (Hint: most of the things you consume are private goods.)
- What are some other examples of public goods?
- How can "free riding" be prevented?
- What are some other examples of externality goods (or externalities in general)?
- Why does it make sense for society to pay for public education, when not everyone has children?

It is understood that a market economy does not fully address all of the needs in a society. These inefficiencies in the way a market economy allocates goods and resources is known as **market failure**. Watch this video to see how externalities and public goods are examples of market failure.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=271#oembed-1

Public goods and externality goods tend to be either supplied too much or supplied too little compared to what is optimal for society. What policies can government pursue to ameliorate this?

Let's find out.

Introduction to Public Goods

Decisions made by firms and individuals in a market often have a spillover effect on other people, whether it be for good (like your neighbor's sweet-smelling cookies), or for the bad (the smelly paper plant that opened up a mile away). In this module, we learn about these positive and negative externalities, as well as public goods, which are typically provided by the government and made freely available to all.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=271

LEARNING OUTCOMES

- Contrast between public and private goods
- Explain the concept of free riders
- Define and give examples of positive and negative externalities
- Analyze the efficacy of government policies to lessen negative externalities and analyze how the government promotes positive externalities
- Analyze the impact of market-based solutions to negative externalities

241. Outcome: Defining Public Goods

What you'll learn to do: contrast between public and private goods

In this outcome, we will define what public goods are and see how the government plays an important role in providing goods for the community large.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Positive Externalities and Public Goods
- Reading: Public Goods
- Video: Public Goods
- Self Check: Defining Public Goods

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

242. Reading: Positive Externalities and Public Goods

Externalities

Private markets, such as the cell phone industry, offer an efficient way to put buyers and sellers together and determine what goods are produced, how they are produced, and who gets them. The principle that voluntary exchange benefits both buyers and sellers is a fundamental building block of the economic way of thinking. But what happens when a voluntary exchange affects a third party who is neither the buyer nor the seller?

As an example, consider a concert producer who wants to build an outdoor arena that will host country music concerts a half-mile from your neighborhood. You will be able to hear these outdoor concerts while sitting on your back porch—or perhaps even in your dining room. In this case, the sellers and buyers of concert tickets may both be quite satisfied with their voluntary exchange, but you have no voice in their market transaction. The effect of a market exchange on a third party who is outside or "external" to the exchange is called an ***externality***. Because externalities that occur in market transactions affect other parties beyond those involved, they are sometimes called ***spillovers***.

Externalities can be negative or positive. If you hate country music, then having it waft into your house every night would be a ***negative externality***. If you love country music, then what amounts to a series of free concerts would be a ***positive externality***.

Positive Externalities in Public Health Programs

One of the most remarkable changes in the standard of living in the last several centuries is that people are living longer. Thousands of years ago, human life expectancy is believed to have been in the range of 20 to 30 years. By 1900, average life expectancy in the United States was 47 years. By the start of the twenty-first century, U.S. life expectancy was 77 years. Most of the gains in life expectancy in the history of the human race happened in the twentieth century.

The rise in life expectancy seems to stem from three primary factors. First, systems for providing clean water and disposing of human waste helped to prevent the transmission of many diseases. Second, changes in public behavior have advanced health. Early in the twentieth century, for example, people learned the importance of boiling bottles before using them for food storage and baby's milk, washing their hands, and protecting food from flies. More recent behavioral changes include reducing the number of people who smoke tobacco and precautions to limit sexually transmitted diseases. Third, medicine has played a large role. Immunizations for diphtheria, cholera, pertussis, tuberculosis, tetanus, and yellow fever were developed between 1890 and 1930. Penicillin, discovered in 1941, led to a series of other antibiotic drugs for bringing infectious diseases under control. In recent decades, drugs that reduce the risks of high blood pressure have had a dramatic effect in extending lives.

These advances in public health have all been closely linked to positive externalities and public goods. Public health officials taught hygienic practices to mothers in the early 1900s and encouraged less smoking in the late 1900s. Many public sanitation systems and storm sewers were funded by government because they have the key traits of public goods. In the twentieth century, many medical discoveries came out of government or university-funded research. Patents and intellectual property rights provided an additional incentive for private inventors. The reason for requiring immunizations, phrased in economic terms, is that it prevents spillovers of illness to others—as well as helping the person immunized.

243. Reading: Public Goods

Public Goods

Consider a situation where the positive externalities are so extensive that private firms could not expect to receive any of the social benefit. This kind of good is called a public good. Spending on national defense is a good example of a public good. Let's begin by defining the characteristics of a public good and discussing why these characteristics make it difficult for private firms to supply public goods. Then we will see how government may step in to address the issue.

The Definition of a Public Good

Economists have a strict definition of a public good, and it does not necessarily include all goods financed through taxes. To understand the defining characteristics of a public good, first consider an ordinary private good, like a piece of pizza. A piece of pizza can be bought and sold fairly easily because it is a separate and identifiable item. However, public goods are not separate and identifiable in this way.

Instead, public goods have two defining characteristics: they are nonexcludable and nonrivalrous. The first characteristic, that a public good is nonexcludable, means that it is costly or impossible to exclude someone from using the good. If Larry buys a private good like a piece of pizza, then he can exclude others, like Lorna, from eating that pizza. However, if national defense is being provided, then it includes everyone. Even if you strongly disagree with America's defense policies or with the level of defense spending, the national defense still protects you. You cannot choose to be unprotected, and national defense cannot protect everyone else and exclude you.

The second main characteristic of a public good, that it is nonrivalrous, means that when one person uses the public good, another can also use it. With a private good like pizza, if Max is eating the pizza then Michelle cannot also eat it; that is, the two people are rivals in consumption. With a public good like national defense, Max's consumption of national defense does not reduce the amount left for Michelle, so they are nonrivalrous in this area.

A number of government services are examples of public goods. For instance, it would not be easy to provide fire and police service so that some people in a neighborhood would be protected from the burning and burglary of their property, while others would not be protected at all. Protecting some necessarily means protecting others, too.

Positive externalities and public goods are closely related concepts. Public goods have positive externalities, like police protection or public health funding. Not all goods and services with positive externalities, however, are public goods. Investments in education have huge positive spillovers but can be provided by a private company. Private companies can invest in new inventions such as the Apple iPad and reap profits that may not capture all of the social benefits. Patents can also be described as an attempt to make new inventions into private goods, which are excludable and rivalrous, so that no one but the inventor is allowed to use them during the length of the patent.

Common Resources and the "Tragedy of the Commons"

There are some goods that do not fall neatly into the categories of private good or public good. While it is easy to classify a pizza as a private good and a city park as a public good, what about an item that is nonexcludable and rivalrous, such as the queen conch?

In the Caribbean, the queen conch is a large marine mollusk found in shallow waters of sea grass. These waters are so shallow, and so clear, that a single diver may harvest many conch in a single day. Not only is conch meat a local delicacy and an important part of the local diet, but the large ornate shells are used in art and can be crafted into musical instruments. Because almost anyone with a small boat, snorkel, and mask, can participate in the conch harvest, it is essentially nonexcludable. At the same time, fishing for conch is rivalrous; once a diver catches one conch it cannot be caught by another diver.

Goods that are nonexcludable and rivalrous are called common resources. Because the waters of the Caribbean are open to all conch fishermen, and because any conch that *you* catch is conch that I cannot catch, common resources like the conch tend to be overharvested.

The problem of overharvesting common resources is not a new one, but ecologist Garret Hardin put the tag “Tragedy of the Commons” to the problem in a 1968 article in the magazine *Science*. Economists view this as a problem of property rights. Since nobody owns the ocean, or the conch that crawl on the sand beneath it, no one individual has an incentive to protect that resource and responsibly harvest it. To address the issue of overharvesting conch and other marine fisheries, economists typically advocate simple devices like fishing licenses, harvest limits, and shorter fishing seasons. When the population of a species drops to critically low numbers, governments have even banned the harvest until biologists determine that the population has returned to sustainable levels. In fact, such is the case with the conch, the harvesting of which has been effectively banned in the United States since 1986.

LINK IT UP

Visit this website for more on the queen conch industry.

Public Goods

Watch this video to review the previous sections about public goods.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=274#oembed-1

Self Check: Defining Public Goods

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/687

244. Outcome: Free Riders

What you'll learn to do: explain the concept of free riders

In this outcome, we will learn about how some public goods lead to the dilemma of not having anyone to pay for the good.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Free Riders
- Self Check: Free Riders

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

245. Reading: Free Riders

Free Riders

Private companies find it difficult to produce public goods. If a good or service is nonexcludable, like national defense, so that it is impossible or very costly to exclude people from using this good or service, then how can a firm charge people for it?

LINK IT UP

Visit this website to read about a connection between free riders and "bad music."

When individuals make decisions about buying a public good, a *free rider* problem can arise, in which people have an incentive to let others pay for the public good and then to "free ride" on the purchases of others. The free rider problem can be expressed in terms of the prisoner's dilemma game, which is discussed as a representation of oligopoly in Monopolistic Competition and Oligopoly. Say that two people are thinking about contributing to a public good: Rachel and Samuel. When either of them contributes to a public good, such as a local fire department, their personal cost of doing so is $4 and the social benefit of that person's contribution is $6. Because society's benefit of $6 is greater than the cost of $4, the investment is a good idea for society as a whole. The problem is that, while Rachel and Samuel pay for the entire cost of their contribution to the public good, they receive only half of the benefit, because the benefit of the public good is divided equally among the members of society. This sets up the prisoner's dilemma illustrated in Table 13.4.

Table 13.4 Contributing to a Public Good as a Prisoner's Dilemma

	Samuel (S) Contribute	Samuel (S) Do Not Contribute
Rachel (R) Contribute	R pays $4, receives $6, net gain +$2 S pays $4, receives $6, net gain +$2	R pays $4, receives $3, net gain −$1 S pays $0, receives $3, net gain +$3
Rachel (R) Do Not Contribute	R pays $0, receives $3, net gain +$3 S pays $4, receives $3, net gain −$1	R pays $0, receives $0 S pays $0, receives $0

If neither Rachel nor Samuel contributes to the public good, then there are no costs and no benefits of the public good. Suppose, however, that only Rachel contributes, while Samuel does not. Rachel incurs a cost of $4, but receives only $3 of benefit (half of the total $6 of benefit to society), while Samuel incurs no cost, and yet he also receives $3 of benefit. In this outcome, Rachel actually loses $1 while Samuel gains $3. A similar outcome, albeit with roles reversed, would occur if Samuel had contributed, but Rachel had not. Finally, if both parties contribute, then each incurs a cost of $4 and each receives $6 of benefit (half of the total $12 benefit to society). There is a dilemma with the Prisoner's Dilemma, though, as you can see in the following example.

THE PROBLEM WITH THE PRISONER'S DILEMMA

The difficulty with the prisoner's dilemma arises as each person thinks through his or her strategic choices.

Step 1. Rachel reasons in this way: If Samuel does not contribute, then I would be a fool to contribute. However, if Samuel does contribute, then I can come out ahead by not contributing.

Step 2. Either way, I should choose not to contribute, and instead hope that I can be a free rider who uses the public good paid for by Samuel.

Step 3. Samuel reasons the same way about Rachel.

Step 4. When both people reason in that way, the public good never gets built, and there is no movement to the option where everyone cooperates—which is actually best for all parties.

THE ROLE OF GOVERNMENT IN PAYING FOR PUBLIC GOODS

The key insight in paying for public goods is to find a way of assuring that everyone will make a contribution and to prevent free riders. For example, if people come together through the political process and agree to pay taxes and make group decisions about the quantity of public goods, they can defeat the free rider problem by requiring, through the law, that everyone contributes.

However, government spending and taxes are not the only way to provide public goods. In some cases, markets can produce public goods. For example, think about radio. It is nonexcludable, since once the radio signal is being broadcast, it would be very difficult to stop someone from receiving it. It is nonrivalrous, since one person listening to the signal does not prevent others from listening as well. Because of these features, it is practically impossible to charge listeners directly for listening to conventional radio broadcasts.

Radio has found a way to collect revenue by selling advertising, which is an indirect way of "charging" listeners by taking up some of their time. Ultimately, consumers who purchase the goods advertised are also paying for the radio service, since the cost of advertising is built into the product cost. In a more recent development, satellite radio companies, such as SiriusXM, charge a regular subscription fee for streaming music without commercials. In this case, however, the product is excludable—only those who pay for the subscription will receive the broadcast.

Some public goods will also have a mixture of public provision at no charge along with fees for some purposes, like a public city park that is free to use, but the government charges a fee for parking your car, for reserving certain picnic grounds, and for food sold at a refreshment stand.

LINK IT UP

Read this article to find out what economists say the government should pay for.

In other cases, social pressures and personal appeals can be used, rather than the force of law, to reduce the number of free riders and to collect resources for the public good. For example, neighbors sometimes form an association to carry out beautification projects or to patrol their area after dark to discourage crime. In low-income countries, where social pressure strongly encourages all farmers to participate, farmers in a region may come together to work on a large irrigation project that will benefit all. Many fundraising efforts, including raising money for local charities and for the endowments of colleges and universities, also can be viewed as an attempt to use social pressure to discourage free riding and to generate the outcome that will produce a public benefit.

Self Check: Free Riders

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/688

246. Outcome: Positive and Negative Externalities

What you'll learn to do: define and give examples of positive and negative externalities

In this section, you will examine the spillover effects of some goods and services—negative effects like pollution and positive effects like better technologies.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Positive Externalities and Technology
- Reading: Introduction to Externalities and Pollution
- Video: Pollution in China
- Video: Externalities
- Reading: Market Failure
- Self Check: Positive and Negative Externalities

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

247. Reading: Positive Externalities and Technology

Positive Externalities and Technology

Market competition can provide an incentive for discovering new technology because a *firm* can earn higher profits by finding a way to produce products more cheaply or to create products with characteristics consumers want. As Gregory Lee, CEO of Samsung said, “Relentless pursuit of new innovation is the key principle of our business and enables consumers to discover a world of possibilities with technology.” An innovative firm knows that it will usually have a temporary edge over its competitors and thus an ability to earn above-normal profits before competitors can catch up.

In certain cases, however, *competition* can discourage new *technology*, especially when other firms can quickly copy a new idea. Consider a pharmaceutical firm deciding to develop a new drug. On average, it can cost $800 million and take more than a decade to discover a new drug, perform the necessary safety tests, and bring the drug to market. If the research and development (R&D) effort fails—and every R&D project has some chance of failure—then the firm will suffer losses and could even be driven out of business. If the project succeeds, then the firm’s competitors may figure out ways of adapting and copying the underlying idea, but without having to pay the costs themselves. As a result, the innovative company will bear the much higher costs of the R&D and will enjoy at best only a small, temporary advantage over the competition.

Many inventors over the years have discovered that their inventions brought them less profit than they might have reasonably expected.

- Eli Whitney (1765–1825) invented the cotton gin, but then southern cotton planters built their own seed-separating devices with a few minor changes in Whitney's design. When Whitney sued, he found that the courts in southern states would not uphold his patent rights.
- Thomas Edison (1847–1931) still holds the record for most patents granted to an individual. His first invention was an automatic vote counter, and despite the social benefits, he could not find a government that wanted to buy it.
- Gordon Gould (1920–2005) came up with the idea behind the laser in 1957. He put off applying for a patent and, by the time he did apply, other scientists had laser inventions of their own. A lengthy legal battle resulted, in which Gould spent $100,000 on lawyers, before he eventually received a patent for the laser in 1977. Compared to the enormous social benefits of the laser, Gould received relatively little financial reward.

A variety of studies by economists have found that the original inventor receives one-third to one-half of the total economic benefits from innovations, while other businesses and new product users receive the rest.

THE POSITIVE EXTERNALITIES OF NEW TECHNOLOGY

Will private firms in a ***market economy*** under invest in research and technology? If a firm builds a factory or buys a piece of equipment, the firm receives all the economic benefits that result from the investments. However, when a firm invests in new technology, the *private benefits*, or profits, that the firm receives are only a portion of the overall social benefits. The *social benefits* of an innovation take into account the value of all the positive externalities of the new idea or product, whether enjoyed by other companies or society as a whole, as well as the private benefits received by the firm that developed the new technology. *Positive externalities* are beneficial spillovers to a third party, or parties.

Consider the example of the Big Drug Company, which is planning its R&D budget for the next year. Economists and scientists working for Big Drug have compiled a list of potential research and development projects and estimated rates of return. (The rate of return is the estimated payoff from the project.) Figure 13.2 shows how the calculations work. The downward-sloping DPrivate curve represents the firm's demand for financial capital and reflects the company's willingness to borrow to finance research and development projects at various interest rates. Suppose that this firm's investment in research and development creates a spillover benefit to other firms and households. After all, new innovations often spark other creative endeavors that society also values. If we add the spillover benefits society enjoys to the firm's private demand for financial capital, we can draw DSocial that lies above DPrivate.

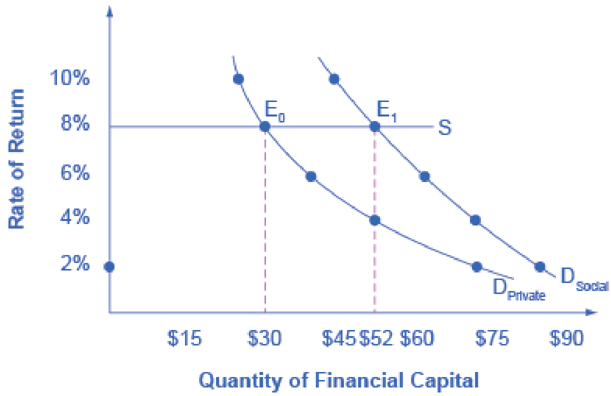


Figure 13.2. *Positive Externalities and Technology.* Big Drug *faces a cost of borrowing of 8%. If the firm receives only the private benefits of investing in R&D, then its demand curve for financial capital is shown by DPrivate, and the equilibrium will occur at $30 million. Because there are spillover benefits, society would find it optimal to have $52 million of investment. If the firm could keep the social benefits of its investment for itself, its demand curve for financial capital would be DSocial and it would be willing to borrow $52 million.*

If there was a way for the firm to fully monopolize those social benefits by somehow making them unavailable to the rest of us, the firm's private demand curve would be the same as society's demand curve. According to Figure 13.2 and Table 13.1, if the going rate of interest on borrowing is 8%, and the company can receive the private benefits of innovation only, then the company would finance $30 million. Society, at the same rate of 8%, would find it optimal to have $52 million of borrowing. Unless there is a way for the company to fully enjoy the total benefits, then it will borrow less than the socially optimal level of $52 million.

Table 13.1. Return and Demand for Capital

Rate of Return	$D_{Private}$ (in millions)	D_{Social} (in millions)
2%	$72	$84
4%	$52	$72
6%	$38	$62
8%	$30	$52
10%	$26	$44

Big Drug's original demand for financial capital (DPrivate) is based on the profits received by the firm. However, other pharmaceutical firms and health care companies may learn new lessons about how to treat certain medical conditions and are then able to create their own competing products. The social benefit of the drug takes into account the value of all the positive externalities of the drug. If Big Drug were able to gain this social return instead of other companies, its *demand* for financial capital would shift to the demand curve DSocial, and it would be willing to borrow and invest $52 million. However, if Big Drug is receiving only 50 cents of each dollar of social benefits, the firm will not spend as much on creating new products. The amount it would be willing to spend would fall somewhere in between DPrivate and DSocial.

WHY INVEST IN HUMAN CAPITAL?

The investment in anything, whether it is the construction of a new power plant or research in a new cancer treatment, usually requires a certain upfront cost with an uncertain future benefit. The investment in education, or human capital, is no different. Over the span of many years, a student and her family invest significant amounts of time and money into that student's education. The idea is that higher levels of educational attainment will eventually serve to increase the student's future productivity and subsequent ability to earn. Once the numbers are crunched, does this investment pay off for the student?

Almost universally, economists have found that the answer to this question is a clear "Yes." For example, several studies of the return to education in the United States estimate that the rate of return to a college education is approximately 10%. Data in Table 13.2, from the U.S. Bureau of Labor Statistics' *Usual Weekly Earnings of Wage and Salary Workers, Third Quarter 2013*, demonstrate that median weekly earnings are higher for workers who have completed more education. While these rates of return will beat equivalent investments in Treasury bonds or savings accounts, the estimated returns to education go primarily to the individual worker, so these returns are ***private rates of return*** to education.

Table 13.2. Usual Weekly Earnings of Wage and Salary Workers, Third Quarter 2013

	Less than a High School Degree	High School Degree, No College	Bachelor's Degree
Median Weekly Earnings (full-time workers over the age of 25)	$479	$659	$1,174

Source: http://www.bls.gov/news.release/pdf/wkyeng.pdf

What does society gain from investing in the education of another student? After all, if the government is spending taxpayer dollars to subsidize public education, society should expect some kind of return on that spending. Again, economists like *George Psacharopoulos* have found that, across a variety of nations, the *social rate of return* on schooling is also positive. After all, positive externalities exist from investment in education. While not always easy to measure, according to *Walter McMahon*, the positive externalities to education typically include better health outcomes for the population, lower levels of crime, a cleaner environment and a more stable, democratic government. For these reasons, many nations have chosen to use taxpayer dollars to subsidize primary, secondary, and higher education. Education clearly benefits the person who receives it, but a society where most people have a good level of education provides positive externalities for all.

OTHER EXAMPLES OF POSITIVE EXTERNALITIES

Although technology may be the most prominent example of a positive externality, it is not the only one. For example, being vaccinated against disease is not only a protection for the individual, but it has the positive spillover of protecting others who may become infected. When a number of homes in a neighborhood are modernized, updated, and restored, not only does it increase the value of those homes, but the value of other properties in the neighborhood may increase as well.

The appropriate public policy response to a positive externality, like a new technology, is to help the party creating the positive externality receive a greater share of the social benefits. In the case of vaccines, like flu shots, an effective policy might be to provide a subsidy to those who choose to get vaccinated.

Figure 13.3 shows the market for flu shots. The market demand curve DMarket for flu shots reflects only the marginal private benefits (MPB) that the vaccinated individuals receive from the shots. Assuming that there are no spillover costs in the production of flu shots, the market supply curve is given by the marginal private cost (MPC) of producing the vaccinations.

The equilibrium quantity of flu shots produced in the market, where MPB is equal to MPC, is QMarket and the price of flu shots is PMarket. However, spillover benefits exist in this market because others, those who chose not to purchase a flu shot, receive a positive externality in a reduced chance of contracting the flu. When we add the spillover benefits to the marginal private benefit of flu shots, the marginal social benefit (MSB) of flu shots is given by DSocial. Because the MPB is greater than MSB, we see that the socially optimal level of flu shots is greater than the market quantity (QSocial exceeds QMarket) and the corresponding price of flu shots, if the market were to produce QSocial, would be at PSocial. Unfortunately, the marketplace does not recognize the positive externality and flu shots will go under produced and under consumed.

So how can government try to move the market level of output closer to the socially desirable level of output? One policy would be to provide a subsidy, like a voucher, to any citizen who wishes to get vaccinated. This voucher would act as "income" that could be used to purchase only a flu shot and, if the voucher was exactly equal to the per-unit spillover benefits, would increase market equilibrium to a quantity of QSocial and a price of PSocial where MSB equals MSC. Suppliers of the flu shots would receive payment of PSocial per vaccination, while consumers of flu shots would redeem the voucher and only pay a price of PSubsidy. When the government uses a subsidy in this way, the socially optimal quantity of vaccinations is produced.

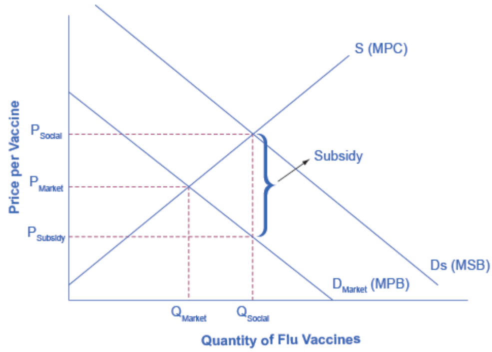


Figure 13.3. *The Market for Flu Shots with Spillover Benefits (A Positive Externality). The market demand curve does not reflect the positive externality of flu vaccinations, so only QMarket will be exchanged. This outcome is inefficient because the marginal social benefit exceeds the marginal social cost. If the government provides a subsidy to consumers of flu shots, equal to the marginal social benefit minus the marginal private benefit, the level of vaccinations can increase to the socially optimal quantity of QSocial.*

248. Reading: Introduction to Externalities and Pollution

Introduction to Externalities and Pollution

In 1969, the Cuyahoga River in Ohio was so polluted that it spontaneously burst into flame. Air pollution was so bad at that time that Chattanooga, Tennessee was a city where, as an article from Sports Illustrated put it: "the death rate from tuberculosis was double that of the rest of Tennessee and triple that of the rest of the United States, a city in which the filth in the air was so bad it melted nylon stockings off women's legs, in which executives kept supplies of clean white shirts in their offices so they could change when a shirt became too gray to be presentable, in which headlights were turned on at high noon because the sun was eclipsed by the gunk in the sky."

Figure 12.1. *Environmental Debate. Across the country, countless people have protested, even risking arrest, against the Keystone XL Pipeline. (Credit: modification of image by "NoKXL"/Flickr Creative Commons)*

The problem of pollution arises for every economy in the world, whether high-income or low-income, and whether market-oriented or command-oriented. Every country needs to strike some balance between production and environmental quality. This module begins by discussing how firms may fail to take certain social costs, like pollution, into their planning if they do not need to pay these costs. Traditionally, policies for environmental protection have focused on governmental limits on how much of each pollutant could be emitted. While this approach has had some success, economists have suggested a range of more flexible, market-oriented policies that reduce pollution at a lower cost. We will consider both approaches, but first let's see how economists frame and analyze these issues.

Keystone XL

You might have heard about Keystone XL in the news. It is a pipeline system designed to bring oil from Canada to the refineries near the Gulf of Mexico, as well as to boost crude oil production in the United States. While a private company, TransCanada, will own the pipeline, U.S. government approval is required because of its size and location. The pipeline is being built in four phases, with the first two currently in operation, bringing oil from Alberta, Canada, east across Canada, south through the United States into Nebraska and Oklahoma, and northeast again to Illinois. The third and fourth phases of the project, known as Keystone XL, would create a pipeline southeast from Alberta straight to Nebraska, and then from Oklahoma to the Gulf of Mexico.

Sounds like a great idea, right? A pipeline that would move much needed crude oil to the Gulf refineries would increase oil production for manufacturing needs, reduce price pressure at the gas pump, and increase overall economic growth. Supporters argue that the pipeline is one of the safest pipelines built yet, and would reduce America's dependence on politically vulnerable Middle Eastern oil imports.

Not so fast, say its critics. The Keystone XL would be constructed over an enormous aquifer (one of the largest in the world) in the Midwest, and through an environmentally fragile area in Nebraska, causing great concern among environmentalists about possible destruction to the natural surroundings. They argue that leaks could taint valuable water sources and construction of the pipeline could disrupt and even harm indigenous species. Environmentalist groups have fought government approval of the proposed construction of the pipeline, and as of press time the pipeline projects remain stalled.

Of course, environmental concerns matter when discussing issues related to economic growth. But how much should they factor in? In the case of the pipeline, how do we know how much damage it would cause when we do not know how to put a value on the environment? Would the benefits of the pipeline outweigh the opportunity cost? The issue of how to balance economic progress with unintended effects on our planet is the subject of this module.

The Economics of Pollution

From 1970 to 2012, the U.S. population increased by one-third and the size of the U.S. economy more than doubled. Since the 1970s, however, the United States, using a variety of anti-pollution policies, has made genuine progress against a number of pollutants. Table 12.1 lists users of energy—from residential to industrial—the types of fuels each used, and the emissions from each, according to the U.S. Energy Information Administration (EIA). The table shows that emissions of certain key air pollutants declined substantially from 2007 to 2012; they dropped 730 million metric tons (MMT) a year—a 12% reduction. This seems to indicate that progress has been made in the United States in reducing overall carbon dioxide emissions, which cause greenhouse gases.

U.S. Carbon Dioxide (CO_2) Emissions from Fossil Fuels Consumed 2007–2012, Million Metric Tons (MMT) per Year

	Primary Fossil Fuels			**Purchased Electric Power**	**Total Primary Fossil Fuels**
End-use Sector	**Coal**	**Petroleum**	**Natural Gas**		
Residential	(0)	(14)	(31)	(134)	(179)
Commercial	(2)	(2)	(7)	(126)	(136)
Industrial	(40)	(62)	31	(118)	(191)
Transportation	0	(228)	5	(1)	(224)
Power	(464)	(36)	(122)	–	–
Change 2007–2012	(508)	(342)	121	(378)	(730)

Despite the gradual reduction in emissions from fossil fuels, many important environmental issues remain. Along with the still high levels of air and water pollution, other issues include hazardous waste disposal, destruction of wetlands and other wildlife habitats, and the impact on human health from pollution.

Pollution as a Negative Externality

Pollution is a negative externality. Economists illustrate the ***social costs*** of production with a demand and supply diagram. The social costs include the private costs of production incurred by the company and the external costs of pollution that are passed on to society. Figure 12.2 shows the demand and supply for manufacturing refrigerators. The demand curve (D) shows the quantity demanded at each price. The supply curve ($S_{private}$) shows the quantity of refrigerators supplied by all the firms at each price if they are taking only their private costs into account and they are allowed to emit pollution at zero cost. The market equilibrium (E_0), where quantity supplied and quantity demanded are equal, is at a price of $650 and a quantity of 45,000. This information is also reflected in the first three columns of Table 12.2.

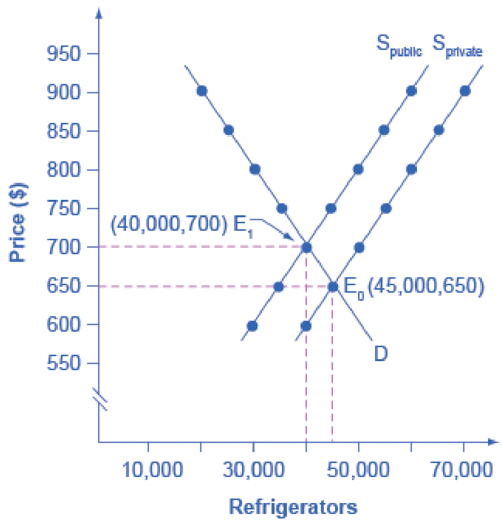


Figure 12.2. *Taking Social Costs into Account: A Supply Shift. If the firm takes only its own costs of production into account, then its supply curve will be Sprivate, and the market equilibrium will occur at E0. Accounting for additional external costs of $100 for every unit produced, the firm's supply curve will be Ssocial. The new equilibrium will occur at E1.*

Table 12.2. A Supply Shift Caused by Pollution Costs Price Quantity Demanded Quantity Supplied before Considering Pollution Cost Quantity Supplied after Considering Pollution Cost $600 50,000

40,000 30,000 $650 45,000 45,000 35,000 $700 40,000 50,000 40,000 $750 35,000 55,000 45,000 $800 30,000 60,000 50,000 $850 25,000 65,000 55,000 $900 20,000 70,000 60,000

However, as a by-product of the metals, plastics, chemicals and energy that are used in manufacturing refrigerators, some pollution is created. Let's say that, if these pollutants were emitted into the air and water, they would create costs of $100 per refrigerator produced. These costs might occur because of injuries to human health, property values, wildlife habitat, reduction of recreation possibilities, or because of other negative impacts. In a market with no anti-pollution restrictions, firms can dispose of certain wastes absolutely free. Now imagine that firms which produce refrigerators must factor in these ***external costs*** of pollution—that is, the firms have to consider not only the costs of labor and materials needed to make a refrigerator, but also the broader costs to society of injuries to health and other values caused by pollution. If the firm is required to pay $100 for the ***additional external costs*** of pollution each time it produces a refrigerator, production becomes more costly and the entire supply curve shifts up by $100.

As illustrated in the fourth column of Table 12.2 and in Figure 12.2, the firm will need to receive a price of $700 per refrigerator and produce a quantity of 40,000—and the firm's new supply curve will be S_{social}. The new equilibrium will occur at E_1, taking the additional external costs of pollution into account results in a higher price, a lower quantity of production, and a lower quantity of pollution. The following feature will walk you through an example, this time with musical accompaniment.

IDENTIFYING THE EQUILIBRIUM PRICE AND QUANTITY

Table 12.3 shows the supply and demand conditions for a firm that will play trumpets on the streets when requested. Output is measured as the number of songs played.

Table 12.3. Supply and Demand Conditions for a Trumpet-Playing Firm

Price	Quantity Demanded	Quantity Supplied without paying the costs of the externality	Quantity Supplied after paying the costs of the externality
$20	0	10	8
$18	1	9	7
$15	2.5	7.5	5.5
$12	4	6	4
$10	5	5	3
$5	7.5	2.5	0.5

Step 1. Determine the negative externality in this situation. To do this, you must think about the situation described and consider all parties that might be impacted. A negative externality might be the increase in noise pollution in the area where the firm is playing.

Step 2. Identify the equilibrium price and quantity when only private costs are taken into account, and then when social costs are taken into account. Remember that equilibrium is where the quantity demanded is equal to the quantity supplied.

Step 3. Look down the columns to where the quantity demanded (the second column) is equal to the "quantity supplied without paying the costs of the externality" (the third column). Then refer to the first column of that row to determine the equilibrium price. In this case, the equilibrium price and quantity when only private costs are taken into account would be at a price of $10 and a quantity of five.

Step 4. Identify the equilibrium price and quantity when the additional external costs are taken into account. Look down the columns of quantity demanded (the second column) and the "quantity supplied after paying the costs of the externality" (the fourth column) then refer to the first column of that row to determine the equilibrium price. In this case, the equilibrium will be at a price of $12 and a quantity of four.

Step 5. Consider how taking the externality into account affects the equilibrium price and quantity. Do this by comparing the two equilibrium situations. If the firm is forced to pay its additional external costs, then production of trumpet songs becomes more costly, and the supply curve will shift up.

Remember that the supply curve is based on choices about production that firms make while looking at their marginal costs, while the demand curve is based on the benefits that individuals perceive while maximizing utility. If no externalities existed, private costs would be the same as the costs to society as a whole, and private benefits would be the same as the benefits to society as a whole. Thus, if no externalities existed, the interaction of demand and supply will coordinate social costs and benefits.

However, when the externality of pollution exists, the supply curve no longer represents all social costs. Because externalities represent a case where markets no longer consider all social costs, but only some of them, economists commonly refer to externalities as an example of ***market failure***. When there is market failure, the private market fails to achieve efficient output, because either firms do not account for all costs incurred in the production of output and/or consumers do not account for all benefits obtained (a positive externality). In the case of pollution, at the market output, social costs of production exceed social benefits to consumers, and the market produces too much of the product.

We can see a general lesson here. If firms were required to pay the social costs of pollution, they would create less pollution but produce less of the product and charge a higher price. In the next module, we will explore how governments require firms to take the social costs of pollution into account.

249. Video: Pollution in China

Pollution in China

Watch this video to see the example of how the entire world is impacted by pollution in China.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=280#oembed-1

250. Video: Externalities

Externalities

Watch the following video to review positive and negative externalities.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=281#oembed-1

251. Reading: Market Failure

Market Failure

Private decisions in the marketplace may not be consistent with the maximization of the net benefit of a particular activity. The failure of private decisions in the marketplace to achieve an efficient allocation of scarce resources is called market failure. Markets will not generate an efficient allocation of resources if they are not competitive or if property rights are not well defined and fully transferable. Either condition will mean that decision makers are not faced with the marginal benefits and costs of their choices.

Imagine you're considering going for an afternoon drive. You'll face some, but not all, of the opportunity costs involved in that choice. In particular, your choice to go for a drive would increase air pollution and might increase traffic congestion. That means that, in weighing the marginal benefits and marginal costs of going for a drive, not all of the costs would be counted. As a result, the net benefit of the allocation of resources such as the air might not be maximized.

External Costs and Benefits

Suppose that in the course of production, the firms in a particular industry generate air pollution. These firms thus impose costs on others, but they do so outside the context of any market exchange—no agreement has been made between the firms and the people affected by the pollution. The firms thus will not be faced with the costs of their action. A cost imposed on others outside of any market exchange is an external cost.

We saw an example of an external cost in our imaginary decision to go for a drive. Here is another: violence on television, in the movies, and in video games. Many critics argue that the violence that pervades these media fosters greater violence in the real world. By the time a child who spends the average amount of time watching television finishes elementary school, he or she will have seen 100,000 acts of violence, including 8,000 murders, according to the American Psychological Association.

Thousands of studies of the relationship between violence in the media and behavior have concluded that there is a link between watching violence and violent behaviors. Video games are a major element of the problem, as young children now spend hours each week playing them. Fifty percent of fourth-grade graders say that their favorite video games are the "first person shooter" type.

Any tendency of increased violence resulting from increased violence in the media constitutes an external cost of such media. The American Academy of Pediatrics reported in 2001 that homicides were the fourth leading cause of death among children between the ages of 10 and 14 and the second leading cause of death for people aged 15 to 24 and has recommended a reduction in exposure to media violence.

An action taken by a person or firm can also create benefits for others, again in the absence of any market agreement; such a benefit is called an external benefit. A firm that builds a beautiful building generates benefits to everyone who admires it; such benefits are external.

External Costs and Efficiency

The case of the polluting firms is illustrated in Figure 6.11 "External Costs." The industry supply curve S_1 reflects private marginal costs, MC_p. The market price is P_p for a quantity Q_p. This is the solution that would occur if firms generating external costs were not forced to pay those costs. If the external costs generated by the pollution were added, the new supply curve S_2 would reflect higher marginal costs, MC_e. Faced with those costs, the market would generate a lower equilibrium quantity, Q_e. That quantity would command a higher price, P_e. The failure to confront producers with the cost of their pollution means that consumers do not pay the full cost of the good they are purchasing.

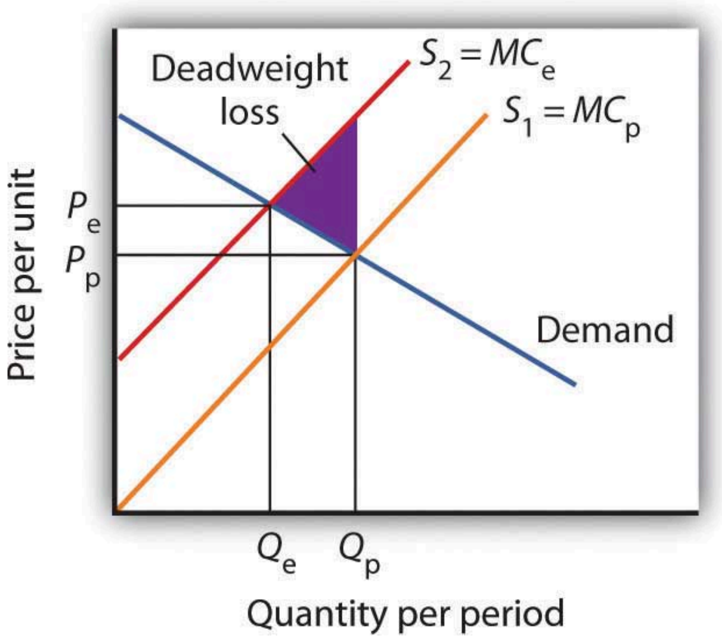


Figure 6.11 External Costs. *When firms in an industry generate external costs, the supply curve S1 reflects only their private marginal costs, MCP. Forcing firms to pay the external costs they impose shifts the supply curve to S2, which reflects the full marginal cost of the firms' production, MCe. Output is reduced and price goes up. The deadweight loss that occurs when firms are not faced with the full costs of their decisions is shown by the shaded area in the graph.*

The level of output and the level of pollution are therefore higher than would be economically efficient. If a way could be found to confront producers with the full cost of their choices, then consumers would be faced with a higher cost as well. Figure 6.11 "External Costs" shows that consumption would be reduced to the efficient level, Q_e, at which demand and the full marginal cost curve (MC_e) intersect. The deadweight loss generated by allowing the external cost to be generated with an output of Q_p is given as the shaded region in the graph.

External Costs and Government Intervention

If an activity generates external costs, the decision makers generating the activity will not be faced with its full costs. Agents who impose these costs will carry out their activities beyond the efficient level; those who consume them, facing too low a price, will consume too much. As a result, producers and consumers will carry out an excessive quantity of the activity.

In such cases, government may try to intervene to reduce the level of the activity toward the efficient quantity. In the case shown in Figure 6.11 "External Costs," for example, firms generating an external cost have a supply curveS_1 that reflects their private marginal costs, MC_p. A per-unit pollution fee imposed on the firms would increase their marginal costs to MC_e, thus shifting the supply curve to S_2, and the efficient level of production would emerge.

Taxes or other restrictions may be imposed on the activity that generates the external cost in an effort to confront decision makers with the costs that they are imposing. In many areas, firms and consumers that pollute rivers and lakes are required to pay fees based on the amount they pollute. Firms in many areas are required to purchase permits in order to pollute the air; the requirement that permits be purchased serves to confront the firms with the costs of their choices.

Another approach to dealing with problems of external costs is direct regulation. For example, a firm may be ordered to reduce its pollution. A person who turns his or her front yard into a garbage dump may be ordered to clean it up. Participants at a raucous party may be told to be quiet. Alternative ways of dealing with external costs are discussed later in the text.

Self Check: Positive and Negative Externalities

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the three Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/689

252. Outcome: Government Involvement and Externalities

What you'll learn to do: analyze the efficacy of government policies to lessen negative externalities and analyze how the government promotes positive externalities

With an understanding of what exactly positive and negative externalities are, we can now examine how the government attempts to promote positive externalities while keeping the negative externalities at bay.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Command-and-Control Regulation
- Reading: The Benefits and Costs of U.S. Environmental Laws
- Reading: How Governments Can Encourage Innovation
- Self Check: Government Involvement and Externalities

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

253. Reading: Command-and-Control Regulation

Command-and-Control Regulation

When the United States started passing comprehensive environmental laws in the late 1960s and early 1970s, a typical law specified how much pollution could be emitted out of a smokestack or a drainpipe and imposed penalties if that limit was exceeded. Other laws required the installation of certain equipment—for example, on automobile tailpipes or on smokestacks—to reduce pollution. These types of laws, which specify allowable quantities of pollution and which also may detail which pollution-control technologies must be used, fall under the category of ***command-and-control regulation***. In effect, command-and-control regulation requires that firms increase their costs by installing anti-pollution equipment; firms are thus required to take the social costs of pollution into account.

Command-and-control regulation has been highly successful in protecting and cleaning up the U.S. environment. In 1970, the Environmental Protection Agency (EPA) was created to oversee all environmental laws. In the same year, the *Clean Air Act* was enacted to address air pollution. Just two years later, in 1972, Congress passed and the president signed the far-reaching *Clean Water Act*. These command-and-control environmental laws, and their amendments and updates, have been largely responsible for America's cleaner air and water in recent decades. However, economists have pointed out three difficulties with command-and-control environmental regulation.

First, command-and-control regulation offers no incentive to improve the quality of the environment beyond the standard set by a particular law. Once the command-and-control regulation has been satisfied, polluters have zero incentive to do better.

Second, command-and-control regulation is inflexible. It usually requires the same standard for all polluters, and often the same pollution-control technology as well. This means that command-and-control regulation draws no distinctions between firms that would find it easy and inexpensive to meet the pollution standard—or to reduce pollution even further—and firms that might find it difficult and costly to meet the standard. Firms have no reason to rethink their production methods in fundamental ways that might reduce pollution even more and at lower cost.

Third, command-and-control regulations are written by legislators and the EPA, and so they are subject to compromises in the political process. Existing firms often argue (and lobby) that stricter environmental standards should not apply to them, only to new firms that wish to start production. Consequently, real-world environmental laws are full of fine print, loopholes, and exceptions.

Although critics accept the goal of reducing pollution, they question whether command-and-control regulation is the best way to design policy tools for accomplishing that goal. A different approach is the use of market-oriented tools, which are discussed in the next section.

254. Reading: The Benefits and Costs of U.S. Environmental Laws

BENEFITS AND COSTS OF CLEAN AIR AND CLEAN WATER

Government economists have estimated that U.S. firms may pay more than $200 billion per year to comply with federal environmental laws. That is big bucks. Is the money well spent?

The benefits of a cleaner environment can be divided into four areas: (1) people may stay healthier and live longer; (2) certain industries that rely on clean air and water, such as farming, fishing, and tourism, may benefit; (3) property values may be higher; and (4) people may simply enjoy a cleaner environment in a way that does not need to involve a market transaction. Some of these benefits, such as gains to tourism or farming, are relatively easy to value in economic terms. It is harder to assign a monetary value to others, such as the value of clean air for someone with asthma. It seems impossible to put a clear-cut monetary value on still others, such as the satisfaction you might feel from knowing that the air is clear over the Grand Canyon, even if you have never visited the Grand Canyon.

Although estimates of environmental benefits are not precise, they can still be revealing. For example, a study by the Environmental Protection Agency looked at the costs and benefits of the Clean Air Act from 1970 to 1990. It found that total costs over that time period were roughly $500 billion—a huge amount. However, it also found that a middle-range estimate of the health and other benefits from cleaner air was $22 trillion—about 44 times higher than the costs. A more recent study by the EPA estimated that the environmental benefits to Americans from the *Clean Air Act* will exceed their costs by a margin of four to one. The EPA estimated that “in 2010 the benefits of Clean Air Act programs will total about $110 billion. This estimate represents the value of avoiding increases in illness and premature death which would have prevailed.” Saying that overall benefits of environmental regulation have exceeded costs in the past, however, is very different from saying that every environmental regulation makes sense. For example, studies suggest that when breaking down emission reductions by type of contaminants, the benefits of air pollution control outweigh the costs primarily for particulates and lead, but when looking at other air pollutants, the costs of reducing them may be comparable to or greater than the benefits. Just because some environmental regulations have had benefits much higher than costs does not prove that every individual regulation is a sensible idea.

ECOTOURISM: MAKING ENVIRONMENTALISM PAY

The definition of ***ecotourism*** is a little vague. Does it mean sleeping on the ground, eating roots, and getting close to wild animals? Does it mean flying in a helicopter to shoot anesthetic darts at African wildlife? Or a little of both? The definition may be fuzzy, but tourists who hope to appreciate the ecology of their destination—”eco tourists”—are the impetus to a big and growing business. The International Ecotourism Society estimates that international tourists interested in seeing nature or wildlife will take 1.56 billion trips by 2020.

LINK IT UP

Visit The International Ecotourism Society's website to learn more about The International Ecotourism Society, its programs, and tourism's role in sustainable community development.

Realizing the attraction of ecotourism, the residents of low-income countries may come to see that preserving wildlife habitats is more lucrative than, say, cutting down forests or grazing livestock to survive. In South Africa, Namibia, and Zimbabwe, for example, a substantial expansion of both rhinoceros and elephant populations is broadly credited to ecotourism, which has given local communities an economic interest in protecting them. Some of the leading ecotourism destinations include: Costa Rica and Panama in Central America; the Caribbean; Malaysia, and other South Pacific destinations; New Zealand; the Serengeti in Tanzania; the Amazon rain forests; and the Galapagos Islands. In many of these countries and regions, governments have enacted policies whereby revenues from ecotourism are shared with local communities, to give people in those local communities a kind of property right that encourages them to conserve their local environment.

Ecotourism needs careful management, so that the combination of eager tourists and local entrepreneurs does not destroy what the visitors are coming to see. But whatever one's qualms are about certain kinds of ecotourism—such as the occasional practice of rich tourists shooting elderly lions with high-powered rifles—it is worth remembering that the alternative is often that low-income people in poor countries will damage their local environment in their effort to survive.

MARGINAL BENEFITS AND MARGINAL COSTS

We can use the tools of marginal analysis to illustrate the marginal costs and the marginal benefits of reducing pollution. Figure 12.4 illustrates a theoretical model of this situation.

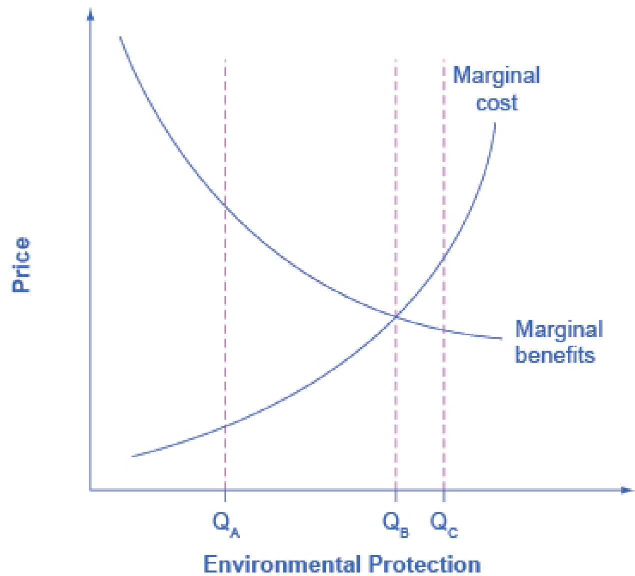


Figure 12.4. *Marginal Costs and Marginal Benefits of Environmental Protection Reducing pollution is costly—resources must be sacrificed. The marginal costs of reducing pollution are generally increasing, because the least expensive and easiest reductions can be made first, leaving the more expensive methods for later. The marginal benefits of reducing pollution are generally declining, because the steps that provide the greatest benefit can be taken first, and steps that provide less benefit can wait until later.*

When the quantity of environmental protection is low so that pollution is extensive—for example, at quantity Qa—there are usually a lot of relatively cheap and easy ways to reduce pollution, and the *marginal benefits* of doing so are quite high. At Qa, it makes sense to allocate more resources to fight pollution. However, as the extent of environmental protection increases, the cheap and easy ways of reducing pollution begin to decrease, and more costly methods must be used. The *marginal cost curve* rises. Also, as environmental protection increases, the largest marginal benefits are achieved first, followed by reduced marginal benefits. As the quantity of environmental protection increases to, say, Qb, the gap between marginal benefits and marginal costs narrows. At point Qc the marginal costs will exceed the marginal benefits. At this level of environmental protection, society is not allocating resources efficiently, because too many resources are being given up to reduce pollution.

As society draws closer to Qb, some might argue that it becomes more important to use market-oriented environmental tools to hold down the costs of reducing pollution. Their objective would be to avoid environmental rules that would provide the quantity of environmental protection at Qc, where marginal costs exceed marginal benefits. The following feature delves into how the EPA measures its policies – and the monetary value of our lives.

WHAT'S A LIFE WORTH?

The U.S. Environmental Protection Agency (EPA) must estimate the value of saving lives by reducing pollution against the additional costs. In measuring the benefits of government environmental policies, the EPA's National Center for Environmental Economics

(NCEE) values a statistical human life at $7.4 million (in 2006 U.S. dollars).

Economists value a human life on the basis of studies of the value that people actually place on human lives in their own decisions. For example, some jobs have a higher probability of death than others, and these jobs typically pay more to compensate for the risk. Examples are ocean fishery as opposed to fish farming, and ice trucking in Alaska as opposed to truck driving in the "lower forty-eight" states.

Government regulators use estimates such as these when deciding what proposed regulations are "reasonable," which means deciding which proposals have high enough benefits to justify their cost. For example, when the U.S. Department of Transportation makes decisions about what safety systems should be required in cars or airplanes, it will approve rules only where the estimated cost per life saved is $3 million or less.

Resources spent on life-saving regulations create tradeoff. A study by W. Kip Viscusi of Vanderbilt University estimated that when a regulation costs $50 million, it diverts enough spending in the rest of the economy from health care and safety expenditures that it costs a life. This finding suggests that any regulation that costs more than $50 million per life saved actually costs lives, rather than saving them.

255. Reading: How Governments Can Encourage Innovation

INTELLECTUAL PROPERTY RIGHTS

A number of different government policies can increase the incentives to innovate, including: guaranteeing intellectual property rights, government assistance with the costs of research and development, and cooperative research ventures between universities and companies.

One way to increase new technology is to guarantee the innovator an exclusive right to that new product or process. *Intellectual property* rights include patents, which give the inventor the exclusive legal right to make, use, or sell the invention for a limited time, and copyright laws, which give the author an exclusive legal right over works of literature, music, film/video, and pictures. For example, if a pharmaceutical firm has a *patent* on a new drug, then no other firm can manufacture or sell that drug for twenty-one years, unless the firm with the patent grants permission. Without a patent, the pharmaceutical firm would have to face competition for any successful products, and could earn no more than a normal rate of profit. With a patent, a firm is able to earn monopoly profits on its product for a period of time—which offers an incentive for research and development. In general, how long can "a period of time" be? The following paragraph discusses patent and copyright protection timeframes for some works you might have heard of.

HOW LONG IS MICKEY MOUSE PROTECTED FROM BEING COPIED?

All patents and copyrights are scheduled to end someday. In 2003, copyright protection for Mickey Mouse was scheduled to run out. Once the copyright had expired, anyone would be able to copy Mickey Mouse cartoons or draw and sell new ones. In 1998, however, Congress passed the Sonny Bono Copyright Term Extension Act. For copyrights owned by companies or other entities, it increased or extended the copyright from 75 years to 95 years after publication. For copyrights owned by individuals, it increased or extended the copyright coverage from 50 years to 70 years after death. Along with protecting Mickey for another 20 years, the copyright extension affected about 400,000 books, movies, and songs.

Figure 13.4 illustrates how the total number of patent applications filed with the U.S. Patent and Trademark Office, as well as the total number of patents granted, surged in the mid-1990s with the invention of the Internet, and is still going strong today.

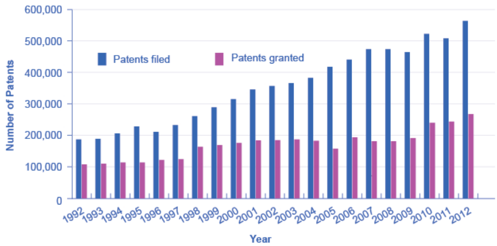


Figure 13.4. *Patents Filed and Granted, 1981–2008 The number of applications filed for patents increased substantially from the mid-1990s into the first years of the 2000s, due in part to the invention of the Internet, which has led to many other inventions and to the 1998 Copyright Term Extension Act. (Source: http://www.uspto.gov/web/offices/ac/ido/oeip/taf/u…)*

While patents provide an incentive to innovate by protecting the innovator, they are not perfect. For example:

- In countries that already have patents, economic studies show that inventors receive only one-third to one-half of the total economic value of their inventions.
- In a fast-moving high-technology industry like biotechnology or semiconductor design, patents may be almost irrelevant

because technology is advancing so quickly.
- Not every new idea can be protected with a patent or a copyright—for example, a new way of organizing a factory or a new way of training employees.
- Patents may sometimes cover too much or be granted too easily. In the early 1970s, Xerox had received over 1,700 patents on various elements of the photocopy machine. Every time Xerox improved the photocopier, it received a patent on the improvement.
- The 21-year time period for a patent is somewhat arbitrary. Ideally, a patent should cover a long enough period of time for the inventor to earn a good return, but not so long that it allows the inventor to charge a monopoly price permanently.

Because patents are imperfect and do not apply well to all situations, alternative methods of improving the rate of return for inventors of new technology are desirable. Some of these possible alternative policies are described in the following sections.

POLICY #1: GOVERNMENT SPENDING ON RESEARCH AND DEVELOPMENT

If the private sector does not have sufficient incentive to carry out research and development, one possibility is for the government to fund such work directly. Government spending can provide direct financial support for research and development (R&D) done at colleges and universities, nonprofit research entities, and sometimes by private firms, as well as at government-run laboratories. While government spending on research and development produces technology that is broadly available for firms to use, it costs taxpayers money and can sometimes be directed more for political than for scientific or economic reasons.

LINK IT UP

Visit the NASA website and the USDA website to read about government research that would not take place where it left to firms due to the externalities.

The first column of Table 13.3 shows the sources of total U.S. spending on research and development; the second column shows the total dollars of R&D funding by each source. The third column shows that, relative to the total amount of funding, 26% comes from the federal government, about 67% of R&D is done by industry, and less than 3% is done by universities and colleges.

Table 13.3 U.S. Research and Development Expenditures, 2008 (Source: http://www.nsf.gov/statistics/nsf12321/pdf/tab10.pdf)

Sources of R&D Funding	Amount ($ billions)	Percent of the Total
Federal government	**$103.7**	**26.08%**
Industry	$267.8	67.35%
Universities and colleges	$10.6	2.67%
Nonprofits	$12	3.02%
Nonfederal government	$3.5	0.88%
Total	$397.6	

In the 1960s the federal government paid for about two-thirds of the nation's R&D. Over time, the U.S. economy has come to rely much more heavily on industry-funded R&D. The federal government has tried to focus its direct R&D spending on areas where private firms are not as active. One difficulty with direct government support of R&D is that it inevitably involves political decisions about which projects are worthy. The scientific question of whether research is worthwhile can easily become entangled with considerations like the location of the congressional district in which the research funding is being spent.

POLICY #2: TAX BREAKS FOR RESEARCH AND DEVELOPMENT

A complementary approach to supporting R&D that does not involve the government's close scrutiny of specific projects is to give firms a reduction in taxes depending on how much research and development they do. The federal government refers to this policy as the research and experimentation (R&E) tax credit. According to the Treasury Department: ". . . the R&E Credit is also a cost-effective policy for stimulating additional private sector investment. Most recent studies find that each dollar of foregone tax revenue through the R&E Tax Credit causes firms to invest at least a dollar in R&D, with some studies finding a benefit to cost ratio of 2 or 2.96."

LINK IT UP

Visit this website for more information on how the R&E Tax Credit encourages investment.

POLICY #3 COOPERATIVE RESEARCH

State and federal governments support research in a variety of ways. For example, United for Medical Research, a coalition of groups that seek funding for the *National Institutes of Health*, (which is supported by federal grants), states: "NIH-supported research added $69 billion to our GDP and supported seven million jobs in 2011 alone." Other institutions, such as the *National Academy of Scientists* and the *National Academy of Engineers*, receive federal grants for innovative projects. The *Agriculture and Food Research Initiative* (*AFRI*) at the United States Department of Agriculture awards federal grants to projects that apply the best science to the most important agricultural problems, from food safety to childhood obesity. Cooperation between government-funded universities, academies, and the private sector can spur product innovation and create whole new industries.

Self Check: Government Involvement and Externalities

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the three Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/690

256. Outcome: Market-Based Solutions

What you'll learn to do: analyze the impact of market-based solutions to negative externalities

You have seen the ways government attempts to encourage positive externalities and discourage the negative, and now you will examine how firms in the market work to combat negative externalities.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Market-Oriented Environmental Tools
- Reading: Market-Oriented Environmental Tools: Effectiveness and Application
- Self Check: Market-Based Solutions

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

257. Reading: Market-Oriented Environmental Tools

Market-Oriented Environmental Tools

Market-oriented environmental policies create incentives to allow firms some flexibility in reducing pollution. The three main categories of market-oriented approaches to pollution control are pollution charges, marketable permits, and better-defined property rights. All of these policy tools, discussed below, address the shortcomings of command-and-control regulation—albeit in different ways.

Pollution Charges

A ***pollution charge*** is a tax imposed on the quantity of pollution that a firm emits. A pollution charge gives a profit-maximizing firm an incentive to figure out ways to reduce its emissions—as long as the marginal cost of reducing the emissions is less than the tax.

For example, consider a small firm that emits 50 pounds per year of small particles, such as soot, into the air. Particulate matter, as it is called, causes respiratory illnesses and also imposes costs on firms and individuals.

Figure 12.3 illustrates the marginal costs that a firm faces in reducing pollution. The marginal cost of pollution reduction, like most most ***marginal cost curves*** increases with output, at least in the short run. Reducing the first 10 pounds of particulate emissions costs the firm $300. Reducing the second 10 pounds would cost $500; reducing the third ten pounds would cost $900; reducing the fourth 10 pounds would cost $1,500; and the fifth 10 pounds would cost $2,500. This pattern for the costs of reducing pollution is common, because the firm can use the cheapest and easiest method to make initial reductions in pollution, but additional reductions in pollution become more expensive.

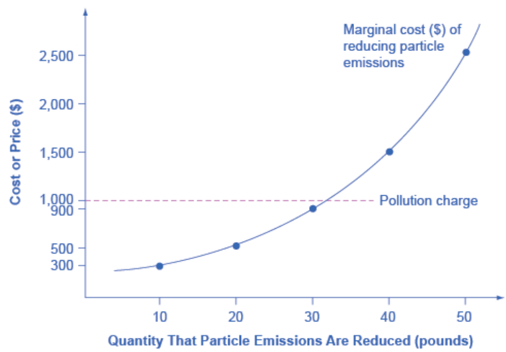


Figure 12.3. *A Pollution Charge If a pollution charge is set equal to $1,000, then the firm will have an incentive to reduce pollution by 30 pounds because the $900 cost of these reductions would be less than the cost of paying the pollution charge.*

Imagine the firm now faces a pollution tax of $1,000 for every 10 pounds of particulates emitted. The firm has the choice of either polluting and paying the tax, or reducing the amount of particulates they emit and paying the cost of abatement as shown in the figure. How much will the firm pollute and how much will the firm abate? The first 10 pounds would cost the firm $300 to abate. This is substantially less than the $1,000 tax, so they will choose to abate. The second 10 pounds would cost $500 to abate, which is still less than the tax, so they will choose to abate. The third 10 pounds would cost $900 to abate, which is slightly less than the $1,000 tax. The fourth 10 pounds would cost $1,500, which is much more costly than paying the tax. As a result, the firm will decide to reduce pollutants by 30 pounds, because the marginal cost of reducing pollution by this amount is less than the pollution tax. With a tax of $1,000, the firm has no incentive to reduce pollution more than 30 pounds.

A firm that has to pay a pollution tax will have an incentive to figure out the least expensive technologies for reducing pollution. Firms that can reduce pollution cheaply and easily will do so to minimize their pollution taxes, whereas firms that will incur high costs for reducing pollution will end up paying the pollution tax instead. If the pollution tax applies to every source of pollution, then no special favoritism or loopholes are created for politically well-connected producers.

For an example of a pollution charge at the household level, consider two ways of charging for garbage collection. One method is to have a flat fee per household, no matter how much garbage a household produces. An alternative approach is to have several levels of fees, depending on how much garbage the household produces—and to offer lower or free charges for recyclable materials. As of 2006 (latest statistics available), the EPA had recorded over 7,000 communities that have implemented "pay as you throw" programs. When people have a financial incentive to put out less garbage and to increase recycling, they find ways of doing so.

Link It Up

Visit this website to learn more about pay-as-you-throw programs, including viewing a map and a table that shows the number of communities using this program in each state.

A number of environmental policies are really pollution charges, although they often do not travel under that name. For example, the federal government and many state governments impose taxes on gasoline. We can view this tax as a charge on the air pollution that cars generate as well as a source of funding for maintaining roads. Indeed, gasoline taxes are far higher in most other countries than in the United States.

Similarly, the refundable charge of five or 10 cents that only 10 states have for returning recyclable cans and bottles works like a pollution tax that provides an incentive to avoid littering or throwing bottles in the trash. Compared with command-and-control regulation, a pollution tax reduces pollution in a more flexible and cost-effective way.

Link It Up

Visit this website to see the current U.S. states with bottle bills and the states that have active campaigns for new bottle bills. You can also view current and proposed bills in Canada and other countries around the world.

Marketable Permits

When a city or state government sets up a ***marketable permit program***, it must start by determining the overall quantity of pollution it will allow as it tries to meet national pollution standards. Then, a number of permits allowing only this quantity of pollution are divided among the firms that emit that pollutant. These permits to pollute can be sold or given to firms free.

Now, add two more conditions. Imagine that these permits are designed to reduce total emissions over time. For example, a permit may allow emission of 10 units of pollution one year, but only nine units the next year, then eight units the year after that, and so on down to some lower level. In addition, imagine that these are marketable permits, meaning that firms can buy and sell them.

To see how marketable permits can work to reduce pollution, consider the four firms listed in Table 12.4. The table shows current emissions of lead from each firm. At the start of the marketable permit program, each firm receives permits to allow this level of pollution. However, these permits are shrinkable, and next year the permits allow the firms to emit only half as much pollution. Let's say that in a year, Firm Gamma finds it easy and cheap to reduce emissions from 600 tons of lead to 200 tons, which means that it has permits that it is not using that allow emitting 100 tons of lead. Firm Beta reduces its lead pollution from 400 tons to 200 tons, so it does not need to buy any permits, and it does not have any extra permits to sell. However, although Firm Alpha can easily reduce pollution from 200 tons to 150 tons, it finds that it is cheaper to purchase permits from Gamma rather than to reduce its own emissions to 100. Meanwhile, Firm Delta did not even exist in the first period, so the only way it can start production is to purchase permits to emit 50 tons of lead.

The total quantity of pollution will decline. But the buying and selling of the marketable permits will determine exactly which firms reduce pollution and by how much. With a system of marketable permits, the firms that find it least expensive to do so will reduce pollution the most.

Table 12.4 How Marketable Permits Work Firm Alpha Firm Beta Firm G

amma Firm Delta Current emissions—permits distributed free for

thisamount200tons400tons600tons0tonsHowmuchpollutionwil

lthesepermitsallowinoneyear?100tons200tons300tons0tonsA

ctualemissionsoneyearinthefuture150tons200tons200tons50

tons Buyer or seller of marketable permit? Buys permits for 50 ton

s Doesn't buy or sell permits Sells permits for 100 tons Buys permi tsfor50tons

Another application of marketable permits occurred when the *Clean Air Act* was amended in 1990. The revised law sought to reduce sulfur dioxide emissions from electric power plants to half of the 1980 levels out of concern that sulfur dioxide was causing acid rain, which harms forests as well as buildings. In this case, the marketable permits the federal government issued were free of charge (no pun intended) to electricity-generating plants across the country, especially those that were burning coal (which produces sulfur dioxide). These permits were of the "shrinkable" type; that is, the amount of pollution allowed by a given permit declined with time.

Better-Defined Property Rights

A clarified and strengthened idea of property rights can also strike a balance between economic activity and pollution. Ronald Coase (1910–2013), who won the 1991 Nobel Prize in economics, offered a vivid illustration of an externality: a railroad track running beside a farmer's field where the railroad locomotive sometimes gives off sparks and sets the field ablaze. Coase asked whose responsibility it was to address this spillover. Should the farmer be required to build a tall fence alongside the field to block the sparks? Or should the railroad be required to put some gadget on the locomotive's smokestack to reduce the number of sparks?

Coase pointed out that this issue cannot be resolved until *property rights* are clearly defined—that is, the legal rights of ownership on which others are not allowed to infringe without paying compensation. Does the farmer have a property right not to have a field burned? Does the railroad have a property right to run its own trains on its own tracks? If neither party has a property right, then the two sides may squabble endlessly, nothing will be done, and sparks will continue to set the field aflame. However, if either the farmer or the railroad has a well-defined legal responsibility, then that party will seek out and pay for the least costly method of reducing the risk that sparks will hit the field. The property right determines whether the farmer or the railroad pays the bills.

The property rights approach is highly relevant in cases involving endangered species. The U.S. government's endangered species list includes about 1,000 plants and animals, and about 90% of these species live on privately owned land. The protection of these endangered species requires careful thinking about incentives and property rights. The discovery of an endangered species on private land has often triggered an automatic reaction from the government to prohibit the landowner from using that land for any purpose that might disturb the imperiled creatures. Consider the incentives of that policy: If you admit to the government that you have an endangered species, the government effectively prohibits you from using your land. As a result, rumors abounded of landowners who followed a policy of "shoot, shovel, and shut up" when they found an endangered animal on their land. Other landowners have deliberately cut trees or managed land in a way that they knew would discourage endangered animals from locating there.

258. Reading: Market-Oriented Environmental Tools: Effectiveness and Application

HOW EFFECTIVE ARE MARKET-ORIENTED ENVIRONMENTAL POLICY TOOLS?

Environmentalists sometimes fear that market-oriented environmental tools are an excuse to weaken or eliminate strict limits on pollution emissions and instead to allow more pollution. It is true that if pollution charges are set very low or if marketable permits do not reduce pollution by very much then market-oriented tools will not work well. But command-and-control environmental laws can also be full of loopholes or have exemptions that do not reduce pollution by much, either. The advantage of market-oriented environmental tools is not that they reduce pollution by more or less, but because of their incentives and flexibility, they can achieve any desired reduction in pollution at a lower cost to society.

A more productive policy would consider how to provide private landowners with an incentive to protect the endangered species that they find and to provide a habitat for additional endangered species. For example, the government might pay landowners who provide and maintain suitable habitats for endangered species or who restrict the use of their land to protect an endangered species. Again, an environmental law built on incentives and flexibility offers greater promise than a command-and-control approach, which tries to oversee millions of acres of privately owned land.

Applying Market-Oriented Environmental Tools

Market-oriented environmental policies are a tool kit. Specific policy tools will work better in some situations than in others. For example, marketable permits work best when a few dozen or a few hundred parties are highly interested in trading, as in the cases of oil refineries that trade lead permits or electrical utilities that trade sulfur dioxide permits. However, for cases in which millions of users emit small amounts of pollution—such as emissions from car engines or unrecycled soda cans—and have no strong interest in trading, pollution charges will typically offer a better choice. Market-oriented environmental tools can also be combined. Marketable permits can be viewed as a form of improved property rights. Or the government could combine marketable permits with a pollution tax on any emissions not covered by a permit.

Self Check: Market-Based Solutions

Answer the question(s) below to see how well you understand the

topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/691

259. Putting It Together: Public Goods

Summary

The goal of this module was to enable you to explain the similarities and differences between public goods and private goods and understand the role for them in the economy.

You learned how to:

- Contrast between public and private goods
- Explain the concept of free riders
- Define and give examples of positive and negative externalities
- Analyze the efficacy of government policies to lessen negative externalities and analyze how the government promotes positive externalities
- Analyze the impact of market-based solutions to negative externalities

Examples

We learned that governments can either promote the supply of positive externality goods or inhibit the supply of negative externality goods through regulation or through financial incentives—subsidies on the former and charges on the latter.

Primary School Children by World Bank Photo Collection, CC-BY-NC-ND.

Let's return to one of the questions from the Why it Matters feature. Why does it make sense for society to pay for public education? It's because education has external benefits—people at large benefit from having an educated population, even beyond the personal benefits they receive from education. Would you rather have educated or ignorant voters? Imagine the benefits a cancer patient would gain if a publically educated scientist discovered the cure, or even new treatment options, for cancer.

260. Glossary: Public Goods

additional external cost
: additional costs incurred by third parties outside the production process when a unit of output is produced

biodiversity
: the full spectrum of animal and plant genetic material

command-and-control regulation
: laws that specify allowable quantities of pollution and that also may detail which pollution-control technologies must be used

externality
: a market exchange that affects a third party who is outside or "external" to the exchange; sometimes called a "spillover"

free rider
: those who want others to pay for the public good and then plan to use the good themselves; if many people act as free riders, the public good may never be provided

intellectual property
: the body of law including patents, trademarks, copyrights, and trade secret law that protect the right of inventors to produce and sell their inventions

international externalities
: externalities that cross national borders and that cannot be resolved by a single nation acting alone

market failure
: When the market on its own does not allocate resources efficiently in a way that balances social costs and benefits; externalities are one example of a market failure

marketable permit program
: a permit that allows a firm to emit a certain amount of pollution; firms with more permits than pollution can sell the remaining permits to other firms

negative externality
: a situation where a third party, outside the transaction, suffers from a market transaction by others

nonexcludable
: when it is costly or impossible to exclude someone from using the good, and thus hard to charge for it

nonrivalrous
: even when one person uses the good, others can also use it

pollution charge
: a tax imposed on the quantity of pollution that a firm emits; also called a pollution tax

positive externalities
: beneficial spillovers to a third party or parties

private benefits
: the dollar value of all benefits of a new product or process invented by a company that can be captured by the investing company

private rates of return
: when the estimated rates of return go primarily to an individual; for example, earning interest on a savings account

property rights
: the legal rights of ownership on which others are not allowed to infringe without paying compensation

public good
: good that is nonexcludable and nonrivalrous, and thus is

difficult for market producers to sell to individual consumers

social benefits
the dollar value of all benefits of a new product or process invented by a company that can be captured by other firms and by society as a whole

social costs
costs that include both the private costs incurred by firms and also additional costs incurred by third parties outside the production process, like costs of pollution

social rate of return
when the estimated rates of return go primarily to society; for example, providing free education

spillover
see externality

261. Discussion: Externalities and Public Goods

Is education a public good? Does education have external benefits? If so, what are some of those public benefits? In your opinion are the external benefits large or small? Why? What should government do to promote the efficient provision of products that have external benefits? If you think that government should put more resources into education, what might be some of the opportunity costs of that decision? How would an economist determine the right amount of government support for education?

PART XV

MODULE: GLOBALIZATION, TRADE AND FINANCE

262. Why It Matters: Globalization, Trade and Finance

Why analyze the benefits and costs of international trade?

This module may be more important than you think. The topic is international trade and includes aspects of globalization and finance, but the theory explains every transaction we conduct. Why do people work for pay instead of growing their own food, building their own house and making their own clothes? Most people are capable of painting their own homes, yet professional painters continue to make a good living. How is international trade different from domestic trade? The answer is "not very much, only in the details." People buy imported goods for the same reasons they buy domestic goods. And yet we often treat foreign and domestic trade as fundamentally different. A grocery chain from a nearby state has recently opened some stores in your neighborhood. How would you feel if the local government prohibited you from shopping at those new stores?

In this module, you will learn that just as buying from the local grocery store is better for most people than growing your own food, so international trade can add to your convenience and quality of life. And yet, most countries support some degree of protectionism, barriers to trade like tariffs, or quotas designed to "protect" domestic workers and companies.

As you proceed through this module, consider the following questions:

- What is comparative advantage?
- What are the gains from international trade?
- In what sense do barriers to trade protect American workers and companies?
- What are the costs of globalization? Are the costs worth it?
- What causes the foreign exchange value of a currency to increase or decrease?
- How does the balance of trade affect the macro economy?

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=294

LEARNING OUTCOMES

- Define and calculate comparative and absolute advantage
- Explain how a nation's workers and consumers are affected by impact of international trade
- Understand the way government regulations (e.g. tariffs, quotas and non-tariff barriers) affect business, consumers and workers in the economy
- Differentiate between alternative international trade regimes and how they impact global trade
- Define currency exchange rates and explain how they influence trade balances
- Explain how the balance of trade (surplus or deficit) affects the domestic economy, and how the domestic economy affects the balance of trade
- Connect globalization, international trade, and international

finance

263. Outcome: Comparative and Absolute Advantage

What you'll learn to do: define and calculate comparative and absolute advantage

In this section, you will learn about the basics behind international trade and why it is helpful for countries to specialize in production of particular goods or services.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Introduction to International Trade
- Reading: Absolute Advantage
- Reading: Absolute and Comparative Advantage
- Reading: Intra-Industry Trade
- Reading: Reducing Barriers to Trade
- Self Check: Comparative and Absolute Advantage

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

264. Reading: Introduction to International Trade

Figure 19.1. *Apple or Samsung iPhone? While the iPhone is readily recognized as an Apple product, 26% of the component costs in it come from components made by rival phone-maker, Samsung. In international trade, there are often "conflicts" like this as each country or company focuses on what it does best. (Credit: modification of work by Yutaka Tsutano Creative Commons)*

JUST WHOSE IPHONE IS IT?

The iPhone is a global product. Apple does not manufacture the iPhone components, nor does it assemble them. The assembly is done by Foxconn Corporation, a Taiwanese company, at its factory in Sengzhen, China. But, Samsung, the electronics firm and competitor to Apple, actually supplies many of the parts that make up an iPhone—about 26%. That means, that Samsung is both the biggest supplier and biggest competitor for Apple. Why do these two firms work together to produce the iPhone? To understand the economic logic behind international trade, you have to accept, as these firms do, that trade is about mutually beneficial exchange. Samsung is one of the world's largest electronics parts suppliers. Apple lets Samsung focus on making the best parts, which allows Apple to concentrate on its strength—designing elegant products that are easy to use. If each company (and by extension each country) focuses on what it does best, there will be gains for all through trade.

We live in a global marketplace. The food on your table might include fresh fruit from Chile, cheese from France, and bottled water from Scotland. Your wireless phone might have been made in Taiwan or Korea. The clothes you wear might be designed in Italy and manufactured in China. The toys you give to a child might have come from India. The car you drive might come from Japan, Germany, or Korea. The gasoline in the tank might be refined from crude oil from Saudi Arabia, Mexico, or Nigeria. As a worker, if your job is involved with farming, machinery, airplanes, cars, scientific instruments, or many other technology-related industries, the odds are good that a hearty proportion of the sales of your employer—and hence the money that pays your salary—comes from export sales. We are all linked by international trade, and the volume of that trade has grown dramatically in the last few decades.

Trade and exchange has been with us since the beginning of human civilization. For example, in the 3rd millennium BCE, southern Mesopotamian communities settled and began to cultivate grains. Agricultural surpluses led to expanded trade with neighboring regions for items not readily available in Mesopotamia such as hardwood, exotic foods and animals, semi-precious stones, hard stones, and metals. This trade transformed subsistence agricultural villages to urban centers and city-states. Scholars, using evidence from the Royal Cemetery at Ur in present day Iraq, suggest that southern Mesopotamian communities traded with other city-states as far as the Indus Valley and Anatolia. Another example of trade before the ascent of Western civilization can be found in China. Chinese maritime power reach its ascendency two centuries before the Europeans during the eighth century. The Ming dynasty used sophisticated maritime equipment unknown in Europe to trade with city states in the far reaches of the Indian Ocean and into the African continent. Similar stories about trade can be told about the Ancient Roman Empire, the Mayans, and the Islamic or Ottoman Empire in 1300 AD.

Despite a long history of international exchange in goods and services, economists such as Paul Krugman argue that the first wave of (western) globalization started in the nineteenth century and and lasted up to the beginning of World War I. Over that time, global exports as a share of global GDP rose from less than 1% of GDP in 1820 to 9% of GDP in 1913. Transportation technology such as steamships and railroads facilitated trade in standardized commodities such as wheat and wool, that were fully global in their reach. When the first submarine telegraph cable was laid under the Atlantic Ocean in 1858, the global flow of information improved and by 1900 all of the world's major economic regions could effectively communicate instantaneously.

Over time the technologies that connect us have changed. They have allowed us to transcend distance faster and to transport goods and services between nations and individuals in different ways. Legal and financial institutions have also changed. With these changes in technologies and institutions, the nature and composition of trade and globalization has changed. Globalization has had positive and negative impacts as we will see in this module. One thing however remains constant: the human instinct to trade comes from a simple insight that trade is about "mutually beneficial exchange."

265. Reading: Absolute and Comparative Advantage

Absolute and Comparative Advantage

The American statesman Benjamin Franklin (1706–1790) once wrote: "No nation was ever ruined by trade." Many economists would express their attitudes toward international trade in an even more positive manner. The evidence that international trade confers overall benefits on economies is pretty strong. Trade has accompanied economic growth in the United States and around the world. Many of the national economies that have shown the most rapid growth in the last few decades—for example, Japan, South Korea, China, and India—have done so by dramatically orienting their economies toward international trade. There is no modern example of a country that has shut itself off from world trade and yet prospered. To understand the benefits of trade, or why we trade in the first place, we need to understand the concepts of comparative and absolute advantage.

In 1817, *David Ricardo*, a businessman, economist, and member of the British Parliament, wrote a treatise called *On the Principles of Political Economy and Taxation*. In this treatise, Ricardo argued that specialization and free trade benefit all trading partners, even those that may be relatively inefficient. To see what he meant, we must be able to distinguish between absolute and comparative advantage.

A country has an *absolute advantage* in producing a good over another country if it uses fewer resources to produce that good. Absolute advantage can be the result of a country's natural endowment. For example, extracting oil in Saudi Arabia is pretty much just a matter of "drilling a hole." Producing oil in other countries can require considerable exploration and costly technologies for drilling and extraction—if indeed they have any oil at all. The United States has some of the richest farmland in the world, making it easier to grow corn and wheat than in many other countries. Guatemala and Colombia have climates especially suited for growing coffee. Chile and Zambia have some of the world's richest copper mines. As some have argued, "geography is destiny." Chile will provide copper and Guatemala will produce coffee, and they will trade. When each country has a product others need and it can be produced with fewer resources in one country over another, then it is easy to imagine all parties benefitting from trade. However, thinking about trade just in terms of geography and absolute advantage is incomplete. Trade really occurs because of comparative advantage.

A country has a *comparative advantage* when a good can be produced at a lower cost in terms of other goods. The question each country or company should be asking when it trades is this: "What do we give up to produce this good?" This should sound like the familiar concept of *opportunity cost* from Choice in a World of Scarcity. For example, if Zambia focuses its resources on producing copper, its labor, land and financial resources cannot be used to produce other goods such as corn. As a result, Zambia gives up the opportunity to produce corn. How do we quantify the cost in terms of other goods? Simplify the problem and assume that Zambia just needs labor to produce copper and corn. The companies that produce either copper or corn tell you that it takes 10 hours to mine a ton of copper and 20 hours to harvest a bushel of corn. This means the opportunity cost of producing a ton of copper is 2 bushels of corn. The next section develops absolute and comparative advantage in greater detail and relates them to trade.

Link It Up

Visit this website for a list of articles and podcasts pertaining to international trade topics.

Watch the following video to better understand why countries benefit from specialization.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=297#oembed-1

A Numerical Example of Absolute and Comparative Advantage

Consider a hypothetical world with two countries, Saudi Arabia and the United States, and two products, oil and corn. Further assume that consumers in both countries desire both these goods. These goods are homogeneous, meaning that consumers/producers cannot differentiate between corn or oil from either country. There is only one resource available in both countries, labor hours. Saudi Arabia can produce oil with fewer resources, while the United States can produce corn with fewer resources. Table 19.1 illustrates the advantages of the two countries, expressed in terms of how many hours it takes to produce one unit of each good.

Table 19.1 How Many Hours It Takes to Produce Oil and Corn

Country	Oil (hours per barrel)	Corn (hours per bushel)
Saudi Arabia	1	4
United States	2	1

In Table 19.1, Saudi Arabia has an absolute advantage in the production of oil because it only takes an hour to produce a barrel of oil compared to two hours in the United States. The United States has an absolute advantage in the production of corn.

To simplify, let's say that Saudi Arabia and the United States each have 100 worker hours (see Table 19.2). We illustrate what each country is capable of producing on its own using a *production possibility frontier* (*PPF*) graph, shown in Figure 19.2. Recall from Choice in a World of Scarcity that the production possibilities frontier shows the maximum amount that each country can produce given its limited resources, in this case workers, and its level of technology.

Table 19.2 Production Possibilities before Trade

Country	Oil Production using 100 worker hours (barrels)		Corn Production using 100 worker hours (bushels)
Saudi Arabia	100	or	25
United States	50	or	100

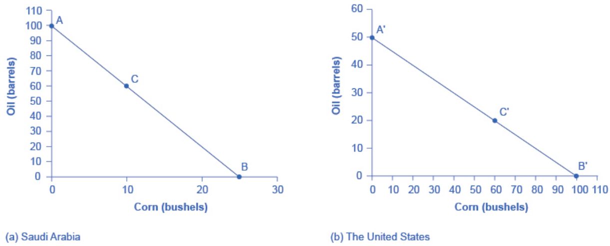


Figure 19.2. *Production Possibilities Frontiers (a) Saudi Arabia can produce 100 barrels of oil at maximum and zero corn (point A), or 25 bushels of corn and zero oil (point B). It can also produce other combinations of oil and corn if it wants to consume both goods, such as at point C. Here it chooses to produce/consume 60 barrels of oil, leaving 40 work hours that can be allocated to producing 10 bushels of corn, using the data in Table 19.1. (b) If the United States produces only oil, it can produce, at maximum, 50 barrels and zero corn (point A'), or at the other*

extreme, it can produce a maximum of 100 bushels of corn and no oil (point B'). Other combinations of both oil and corn are possible, such as point C'. All points above the frontiers are impossible to produce given the current level of resources and technology.

Arguably Saudi and U.S. consumers desire both oil and corn to live. Let's say that before trade occurs, both countries produce and consume at point C or C'. Thus, before trade, the Saudi Arabian economy will devote 60 worker hours to produce oil, as shown in Table 19.3. Given the information in Table 19.1, this choice implies that it produces/consumes 60 barrels of oil. With the remaining 40 worker hours, since it needs four hours to produce a bushel of corn, it can produce only 10 bushels. To be at point C', the U.S. economy devotes 40 worker hours to produce 20 barrels of oil and the remaining worker hours can be allocated to produce 60 bushels of corn.

Production before Trade

Country	**Oil Production (barrels)**	**Corn Production (bushels)**
Saudi Arabia (C)	60	10
United States (C')	20	60
Total World Production	**80**	**70**

The slope of the production possibility frontier illustrates the opportunity cost of producing oil in terms of corn. Using all its resources, the United States can produce 50 barrels of oil *or* 100 bushels of corn. So the opportunity cost of one barrel of oil is two bushels of corn—or the slope is 1/2. Thus, in the U.S. production possibility frontier graph, every increase in oil production of one barrel implies a decrease of two bushels of corn. Saudi Arabia can produce 100 barrels of oil *or* 25 bushels of corn. The opportunity cost of producing one barrel of oil is the loss of 1/4 of a bushel of corn that Saudi workers could otherwise have produced. In terms of corn, notice that Saudi Arabia gives up the least to produce a barrel of oil. These calculations are summarized in Table 19.4.

Opportunity Cost and Comparative Advantage

Country	Opportunity cost of one unit — Oil (in terms of corn)	Opportunity cost of one unit — Corn (in terms of oil)
Saudi Arabia	¼	4
United States	2	½

Again recall that comparative advantage was defined as the opportunity cost of producing goods. Since Saudi Arabia gives up the least to produce a barrel of oil, (¼ < 2 in Table 19.4) it has a comparative advantage in oil production. The United States gives up the least to produce a bushel of corn, so it has a comparative advantage in corn production.

In this example, there is symmetry between absolute and comparative advantage. Saudi Arabia needs fewer worker hours to produce oil (absolute advantage, see Table 19.1), and also gives up the least in terms of other goods to produce oil (comparative advantage, see Table 19.4). Such symmetry is not always the case, as we will show after we have discussed gains from trade fully. But first, read the following feature to make sure you understand why the PPF line in the graphs is straight.

Can a production possibility frontier be straight?

When you first met the production possibility frontier (PPF) in the module on Choice in a World of Scarcity it was drawn with an outward-bending shape. This shape illustrated that as inputs were transferred from producing one good to another—like from education to health services—there were increasing opportunity costs. In the examples in this module, the PPFs are drawn as straight lines, which means that opportunity costs are constant. When a marginal unit of labor is transferred away from growing corn and toward producing oil, the decline in the quantity of corn and the increase in the quantity of oil is always the same. In reality this is possible only if the contribution of additional workers to output did not change as the scale of production changed. The linear production possibilities frontier is a less realistic model, but a straight line simplifies calculations. It also illustrates economic themes like absolute and comparative advantage just as clearly.

266. Reading: Absolute Advantage

What Happens When a Country Has an Absolute Advantage in All Goods

What happens to the possibilities for trade if one country has an absolute advantage in everything? This is typical for high-income countries that often have well-educated workers, technologically advanced equipment, and the most up-to-date production processes. These high-income countries can produce all products with fewer resources than a low-income country. If the high-income country is more productive across the board, will there still be gains from trade? Good students of Ricardo understand that trade is about mutually beneficial exchange. Even when one country has an absolute advantage in all products, trade can still benefit both sides. This is because gains from trade come from specializing in one's comparative advantage.

Production Possibilities and Comparative Advantage

Consider the example of trade between the United States and Mexico described in Table 19.1. In this example, it takes four U.S. workers to produce 1,000 pairs of shoes, but it takes five Mexican workers to do so. It takes one U.S. worker to produce 1,000 refrigerators, but it takes four Mexican workers to do so. The United States has an absolute advantage in productivity with regard to both shoes and refrigerators; that is, it takes fewer workers in the United

States than in Mexico to produce both a given number of shoes and a given number of refrigerators.

Table 19.1. Resources Needed to Produce Shoes and Refrigerators

Country	Number of Workers needed to produce 1,000 units — Shoes	Number of Workers needed to produce 1,000 units — Refrigerators
United States	4 workers	1 worker
Mexico	5 workers	4 workers

Absolute advantage simply compares the *productivity* of a worker between countries. It answers the question, "How many inputs do I need to produce shoes in Mexico?" Comparative advantage asks this same question slightly differently. Instead of comparing how many workers it takes to produce a good, it asks, "How much am I giving up to produce this good in this country?" Another way of looking at this is that comparative advantage identifies the good for which the producer's absolute advantage is relatively larger, or where the producer's absolute productivity disadvantage is relatively smaller. The United States can produce 1,000 shoes with four-fifths as many workers as Mexico (four versus five), but it can produce 1,000 refrigerators with only one-quarter as many workers (one versus four). So, the comparative advantage of the United States, where its absolute productivity advantage is relatively greatest, lies with refrigerators, and Mexico's comparative advantage, where its absolute productivity disadvantage is least, is in the production of shoes.

Mutually Beneficial Trade with Comparative Advantage

When nations increase production in their area of comparative advantage and trade with each other, both countries can benefit.

Again, the production possibility frontier is a useful tool to visualize this benefit.

Consider a situation where the United States and Mexico each have 40 workers. For example, as Table 19.2 shows, if the United States divides its labor so that 40 workers are making shoes, then, since it takes four workers in the United States to make 1,000 shoes, a total of 10,000 shoes will be produced. (If four workers can make 1,000 shoes, then 40 workers will make 10,000 shoes). If the 40 workers in the United States are making refrigerators, and each worker can produce 1,000 refrigerators, then a total of 40,000 refrigerators will be produced.

Table 19.2. Production Possibilities before Trade with Complete Specialization

Country	**Shoe Production — using 40 workers**		**Refrigerator Production — using 40 workers**
United States	10,000 shoes	or	40,000 refrigerators
Mexico	8,000 shoes	or	10,000 refrigerators

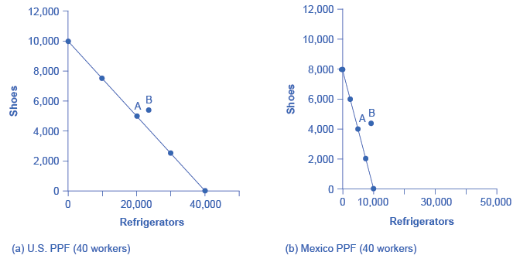


Figure 19.1. *Production Possibility Frontiers. (a) With 40 workers, the United States can produce either 10,000 shoes and zero refrigerators or 40,000 refrigerators and zero shoes. (b) With 40 workers, Mexico can produce a maximum of 8,000 shoes and zero refrigerators, or 10,000 refrigerators and zero shoes. All other points on the production possibility line are possible combination s of the two goods that can be produced given current resources. Point A on both graphs is where the countries start producing and consuming*

before trade. Point B is where they end up after trade.

As always, the slope of the production possibility frontier for each country is the opportunity costs as labor is transferred from shoe production to refrigerators, or vice versa (see Figure 19.1).

Let's say that, in the situation before trade, each nation prefers to produce a combination of shoes and refrigerators that is shown at point A. Table 19.3 shows the output of each good for each country and the total output for the two countries.

Table 19.3 Total Production at Point A before Trade

Country	**Current Shoe Production**	**Current Refrigerator Production**
United States	5,000	20,000
Mexico	4,000	5,000
Total	**9,000**	**25,000**

Continuing with this scenario, each country transfers some amount of labor toward its area of comparative advantage. For example, the United States transfers six workers away from shoes and toward producing refrigerators. As a result, U.S. production of shoes decreases by 1,500 units (6/4 × 1,000), while its production of refrigerators increases by 6,000 (that is, 6/1 × 1,000). Mexico also moves production toward its area of comparative advantage, transferring 10 workers away from refrigerators and toward production of shoes. As a result, production of refrigerators in Mexico falls by 2,500 (10/4 × 1,000), but production of shoes increases by 2,000 pairs (10/5 × 1,000). Notice that when both countries shift production toward each of their comparative advantages (what they are relatively better at), their combined production of both goods rises, as shown in Table 19.4. The reduction of shoe production by 1,500 pairs in the United States is more than offset by the gain of 2,000 pairs of shoes in Mexico, while the reduction of 2,500 refrigerators in Mexico is more than offset by the additional 6,000 refrigerators produced in the United States.

Table 19.4 Shifting Production Toward Comparative Advantage Raises Total Output

Country	Shoe Production	Refrigerator Production
United States	3,500	26,000
Mexico	6,000	2,500
Total	**9,500**	**28,500**

This numerical example illustrates the remarkable insight of comparative advantage: even when one country has an absolute advantage in all goods and another country has an absolute disadvantage in all goods, both countries can still benefit from trade. Even though the United States has an absolute advantage in producing both refrigerators and shoes, it makes economic sense for it to specialize in the good for which it has a comparative advantage. The United States will export refrigerators and in return import shoes.

How Opportunity Cost Sets the Boundaries of Trade

This example shows that both parties can benefit from specializing in their comparative advantages and trading. By using the opportunity costs in this example, it is possible to identify the range of possible trades that would benefit each country.

Mexico started out, before specialization and trade, producing 4,000 pairs of shoes and 5,000 refrigerators (see Figure 19.1 and Table 19.3). Then, in the numerical example given, Mexico shifted production toward its comparative advantage and produced 6,000 pairs of shoes but only 2,500 refrigerators. Thus, if Mexico can *export* no more than 2,000 pairs of shoes (giving up 2,000 pairs of shoes) in exchange for *imports* of at least 2,500 refrigerators (a gain of 2,500 refrigerators), it will be able to consume more of both goods than before trade. Mexico will be unambiguously better off. Conversely, the United States started off, before specialization and trade, producing 5,000 pairs of shoes and 20,000 refrigerators. In the example, it then shifted production toward its comparative advantage, producing only 3,500 shoes but 26,000 refrigerators. If the United States can export no more than 6,000 refrigerators in exchange for imports of at least 1,500 pairs of shoes, it will be able to consume more of both goods and will be unambiguously better off.

The range of trades that can benefit both nations is shown in Table 19.5. For example, a trade where the U.S. exports 4,000 refrigerators to Mexico in exchange for 1,800 pairs of shoes would benefit both sides, in the sense that both countries would be able to consume more of both goods than in a world without trade.

Table 19.5. The Range of Trades That Benefit Both the United States and Mexico

The U.S. economy, after specialization, will benefit if it:	The Mexican economy, after specialization, will benefit if it:
Exports fewer than 6,000 refrigerators	*Imports* at least 2,500 refrigerators
Imports at least 1,500 pairs of shoes	*Exports* no more than 2,000 pairs of shoes

Trade allows each country to take advantage of lower opportunity costs in the other country. If Mexico wants to produce more refrigerators without trade, it must face its domestic opportunity costs and reduce shoe production. If Mexico, instead, produces more shoes and then trades for refrigerators made in the United States, where the *opportunity cost* of producing refrigerators is lower, Mexico can in effect take advantage of the lower opportunity cost of refrigerators in the United States. Conversely, when the United States specializes in its comparative advantage of refrigerator production and trades for shoes produced in Mexico, international trade allows the United States to take advantage of the lower opportunity cost of shoe production in Mexico.

The theory of comparative advantage explains why countries trade: they have different comparative advantages. It shows that the gains from international trade result from pursuing comparative advantage and producing at a lower opportunity cost. The following feature shows how to calculate absolute and comparative advantage and the way to apply them to a country's production.

Calculating Absolute and Comparative Advantage

In Canada a worker can produce 20 barrels of oil or 40 tons of lumber. In Venezuela, a worker can produce 60 barrels of oil or 30 tons of lumber.

Table 19.6

Country	Oil(barrels)		Lumber(tons)
Canada	20	or	40
Venezuela	60	or	30

1. Who has the absolute advantage in the production of oil or lumber? How can you tell?
2. Which country has a comparative advantage in the production of oil?
3. Which country has a comparative advantage in producing lumber?
4. In this example, is absolute advantage the same as comparative advantage, or not?
5. In what product should Canada specialize? In what product should Venezuela specialize?

Step 1. Make a table like Table 19.6.

Step 2. To calculate absolute advantage, look at the larger of the numbers for each product. One worker in Canada can produce more lumber (40 tons versus 30 tons), so Canada has the absolute advantage in lumber. One worker in Venezuela can produce 60 barrels of oil compared to a worker in Canada who can produce only 20.

Step 3. To calculate comparative advantage, find the opportunity cost of producing one barrel of oil in both countries. The country with the lowest opportunity cost has the comparative advantage. With the same labor time, Canada can produce either 20 barrels of oil or 40 tons of lumber. So in effect, 20 barrels of oil is equivalent to 40 tons of lumber: 20 oil = 40 lumber. Divide both sides of the equation by 20 to calculate the opportunity cost of one barrel of oil in Canada. 20/20 oil = 40/20 lumber. 1 oil = 2 lumber. To produce one additional barrel of oil in Canada has an opportunity cost of 2 lumber. Calculate the same way for Venezuela: 60 oil = 30 lumber. Divide both sides of the equation by 60. One oil in Venezuela has an opportunity cost of 1/3 lumber. Because 1/3 lumber < 2 lumber, Venezuela has the comparative advantage in producing oil.

Step 4. Calculate the opportunity cost of one lumber by reversing the numbers, with lumber on the left side of the equation. In Canada, 40 lumber is equivalent in labor time to 20 barrels of oil: 40 lumber = 20 oil. Divide each side of the equation by 40. The opportunity cost of one lumber is 1/2 oil. In Venezuela, the equivalent labor time will produce 30 lumber or 60 oil: 30 lumber = 60 oil. Divide each side by 30. One lumber has an opportunity cost of two oil. Canada has the lower opportunity cost in producing lumber.

Step 5. In this example, absolute advantage is the same as comparative advantage. Canada has the absolute and comparative advantage in lumber; Venezuela has the absolute and comparative advantage in oil.

Step 6. Canada should specialize in what it has a relative lower opportunity cost, which is lumber, and Venezuela should specialize in oil. Canada will be exporting lumber and importing oil, and Venezuela will be exporting oil and importing lumber.

Comparative Advantage Goes Camping

To build an intuitive understanding of how comparative advantage can benefit all parties, set aside examples that involve national economies for a moment and consider the situation of a group of friends who decide to go camping together. The six friends have a wide range of skills and experiences, but one person in particular, Jethro, has done lots of camping before and is also a great athlete. Jethro has an absolute advantage in all aspects of camping: he is faster at carrying a backpack, gathering firewood, paddling a canoe, setting up tents, making a meal, and washing up. So here is the question: Because Jethro has an absolute productivity advantage in everything, should he do all the work?

Of course not! Even if Jethro is willing to work like a mule while everyone else sits around, he, like most mortals, only has 24 hours in a day. If everyone sits around and waits for Jethro to do everything, not only will Jethro be an unhappy camper, but there will not be much output for his group of six friends to consume. The theory of comparative advantage suggests that everyone will benefit if they figure out their areas of comparative advantage—that is, the area of camping where their productivity disadvantage is least, compared to Jethro. For example, it may be that Jethro is 80% faster at building fires and cooking meals than anyone else, but only 20% faster at gathering firewood and 10% faster at setting up tents. In that case, Jethro should focus on building fires and making meals, and others should attend to the other tasks, each according to where their productivity disadvantage is smallest. If the campers coordinate their efforts according to comparative advantage, they can all gain.

267. Reading: Intra-Industry Trade

The Prevalence of Intra-industry Trade between Similar Economies

Absolute and comparative advantages explain a great deal about patterns of global trade. For example, they help to explain the patterns noted at the start of this module, like why you may be eating fresh fruit from Chile or Mexico, or why lower productivity regions like Africa and Latin America are able to sell a substantial proportion of their exports to higher productivity regions like the European Union and North America. Comparative advantage, however, at least at first glance, does not seem especially well-suited to explain other common patterns of international trade.

The theory of comparative advantage suggests that trade should happen between economies with large differences in opportunity costs of production. Roughly half of all world trade involves shipping goods between the fairly similar high-income economies of the United States, Canada, the European Union, Japan, Mexico, and China (see Table 19.7).

Table 19.7. Where U.S. Exports Go and U.S. Imports Originate (2013)		
Country	**U.S. Exports Go to . . .**	**U.S. Imports Come from . . .**
European Union	17.0%	16.0%
Canada	19.0%	14.0%
Japan	5.0%	6.0%
Mexico	14.0%	12.0%
China	7.0%	18.0%
Source: http://www.census.gov/foreign-trade/statistics/highlights/toppartners.html		

Moreover, the theory of comparative advantage suggests that each economy should specialize to a degree in certain products, and then exchange those products. A high proportion of trade, however, is *intra-industry trade*—that is, trade of goods within the same industry from one country to another. For example, the United States produces and exports autos and imports autos. Table 19.8 shows some of the largest categories of U.S. exports and imports. In all of these categories, the United States is both a substantial exporter and a substantial importer of goods from the same industry. In 2012, according to the Bureau of Economic Analysis, the United States exported $146 billion worth of autos, and imported $298 billion worth of autos. About 60% of U.S. trade and 60% of European trade is intra-industry trade.

Table 19.8 Some Intra-Industry U.S. Exports and Imports in 2012

Some U.S. Exports	Quantity of Exports ($ billions)	Quantity of Imports ($ billions)
Autos	$146	$298
Food and beverages	$133	$110
Capital goods	$527	$549
Consumer goods	$182	$516
Passenger fares	$39	$35
Other transportation	$45	$55

Source: http://bea.gov/newsreleases/international/trade/2013/pdf/trad1212.pdf

Why do similar high-income economies engage in intra-industry trade? What can be the economic benefit of having workers of fairly similar skills making cars, computers, machinery and other products which are then shipped across the oceans to and from the United States, the European Union, and Japan? There are two reasons: (1) The *division of labor* leads to learning, innovation, and unique skills; and (2) economies of scale.

Gains from Specialization and Learning

Consider the category of machinery, where the U.S. economy has considerable intra-industry trade. Machinery comes in many varieties, so the United States may be exporting machinery for manufacturing with wood, but importing machinery for photographic processing. The underlying reason why a country like the United States, Japan, or Germany produces one kind of machinery rather than another is usually not related to U.S., German, or Japanese firms and workers having generally higher or lower skills. It is just that, in working on very specific and particular products, firms in certain countries develop unique and different skills.

Specialization in the world economy can be very finely split. In fact, recent years have seen a trend in international trade called *splitting up the value chain*. The *value chain* describes how a good is produced in stages. As indicated in the beginning of the module, the production of the iPhone involves the design and engineering of the phone in the United States, parts supplied from Korea, the assembly of the parts in China, and the advertising and marketing done in the United States. Thanks in large part to improvements in communication technology, sharing information, and transportation, it has become easier to split up the value chain.

Instead of production in a single large factory, all of these steps can be split up among different firms operating in different places and even different countries. Because firms split up the value chain, international trade often does not involve whole finished products like automobiles or refrigerators being traded between nations. Instead, it involves shipping more specialized goods like, say, automobile dashboards or the shelving that fits inside refrigerators. Intra-industry trade between similar countries produces economic gains because it allows workers and firms to learn and innovate on particular products—and often to focus on very particular parts of the value chain.

Link It Up

Visit this website for some interesting information about the assembly of the iPhone.

Economies of Scale, Competition, Variety

A second broad reason that intra-industry trade between similar nations produces economic gains involves economies of scale. The concept of *economies of scale* means that as the scale of output goes up, average costs of production decline—at least up to a point. Figure 19.2 illustrates economies of scale for a plant producing toaster ovens. The horizontal axis of the figure shows the quantity of production by a certain firm or at a certain manufacturing plant. The vertical axis measures the average cost of production. Production plant S produces a small level of output at 30 units and has an average cost of production of $30 per toaster oven. Plant M produces at a medium level of output at 50 units, and has an average cost of production of $20 per toaster oven. Plant L produces 150 units of output with an average cost of production of only $10 per toaster oven. Although plant V can produce 200 units of output, it still has the same unit cost as Plant L.

In this example, a small or medium plant, like S or M, will not be able to compete in the market with a large or a very large plant like L or V, because the firm that operates L or V will be able to produce and sell their output at a lower price. In this example, economies of scale operate up to point L, but beyond point L to V, the additional scale of production does not continue to reduce average costs of production.

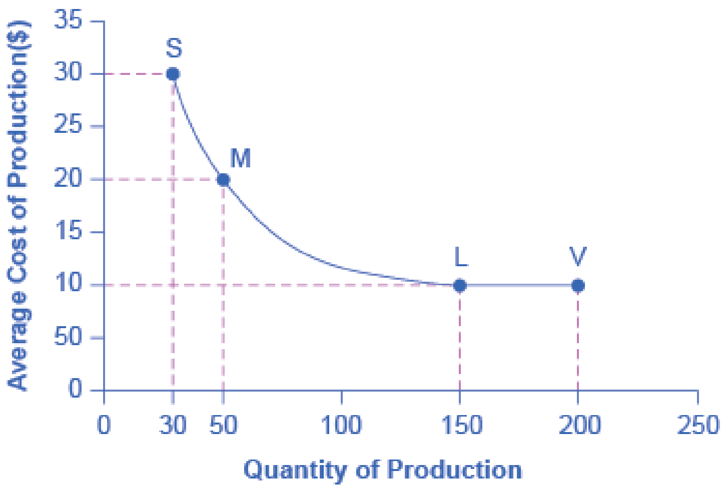


Figure 19.2 *Economies of Scale Production. Plant S, has an average cost of production of $30 per toaster oven. Production plant M has an average cost of production of $20 per toaster oven. Production plant L has an average cost of production of only $10 per toaster oven. Production plant V would still have an average cost of production of $10 per toaster oven. Thus, production plant M can produce toaster ovens more cheaply than plant S because of economies of scale, and plants L or V can produce more cheaply than S or M because of economies of scale. However, the economies of scale end at an output level of 150. Plant V, despite being larger, cannot produce more cheaply on average than plant L.*

The concept of economies of scale becomes especially relevant to international trade when it enables one or two large producers to supply the entire country. For example, a single large automobile factory could probably supply all the cars purchased in a smaller economy like the United Kingdom or Belgium in a given year. However, if a country has only one or two large factories producing cars, and no *international trade*, then consumers in that country would have relatively little choice between kinds of cars (other than the color of the paint and other nonessential options). Little or no competition will exist between different car manufacturers.

International trade provides a way to combine the lower average production costs that come from economies of scale and still have competition and variety for consumers. Large automobile factories in different countries can make and sell their products around the world. If the U.S. automobile market was made up of only General Motors, Ford, and Chrysler, the level of competition and consumer choice would be quite a lot lower than when U.S. carmakers must face competition from Toyota, Honda, Suzuki, Fiat, Mitsubishi, Nissan, Volkswagen, Kia, Hyundai, BMW, Subaru, and others. Greater competition brings with it innovation and responsiveness to what consumers want. America's car producers make far better cars now than they did several decades ago, and much of the reason is competitive pressure, especially from East Asian and European carmakers.

Dynamic Comparative Advantage

The sources of gains from intra-industry trade between similar economies—namely, the learning that comes from a high degree of specialization and splitting up the value chain and from economies of scale—do not contradict the earlier theory of comparative advantage. Instead, they help to broaden the concept.

In intra-industry trade, the level of worker productivity is not determined by climate or geography. It is not even determined by the general level of education or skill. Instead, the level of worker productivity is determined by how firms engage in specific learning about specialized products, including taking advantage of economies of scale. In this vision, comparative advantage can be dynamic—that is, it can evolve and change over time as new skills are developed and as the value chain is split up in new ways. This line of thinking also suggests that countries are not destined to have the same comparative advantage forever, but must instead be flexible in response to ongoing changes in comparative advantage.

268. Reading: Reducing Barriers to Trade

Reducing Barriers to Trade

Tariffs are taxes that governments place on imported goods for a variety of reasons. Some of these reasons include protecting sensitive industries, for humanitarian reasons, and protecting against *dumping*. Traditionally, tariffs were used simply as a political tool to protect certain vested economic, social, and cultural interests. The *World Trade Organization* (WTO) is committed to lowering barriers to trade. The world's nations meet through the WTO to negotiate how they can reduce barriers to trade, such as tariffs. WTO negotiations happen in "rounds," where all countries negotiate one agreement to encourage trade, take a year or two off, and then start negotiating a new agreement. The current round of negotiations is called the Doha Round because it was officially launched in Doha, the capital city of Qatar, in November 2001. In 2009, economists from the World Bank summarized recent research and found that the Doha round of negotiations would increase the size of the world economy by $160 billion to $385 billion per year, depending on the precise deal that ended up being negotiated.

In the context of a global economy that currently produces more than $30 trillion of goods and services each year, this amount is not huge: it is an increase of 1% or less. But before dismissing the gains from trade too quickly, it is worth remembering two points.

- First, a gain of a few hundred billion dollars is enough money to deserve attention! Moreover, remember that this increase is not a one-time event; it would persist each year into the future.

- Second, the estimate of gains may be on the low side because some of the gains from trade are not measured especially well in economic statistics. For example, it is difficult to measure the potential advantages to consumers of having a variety of products available and a greater degree of competition among producers. Perhaps the most important unmeasured factor is that trade between countries, especially when firms are splitting up the value chain of production, often involves a transfer of knowledge that can involve skills in production, technology, management, finance, and law.

Low-income countries benefit more from trade than high-income countries do. In some ways, the giant U.S. economy has less need for international trade, because it can already take advantage of internal trade within its economy. However, many smaller national economies around the world, in regions like Latin America, Africa, the Middle East, and Asia, have much more limited possibilities for trade inside their countries or their immediate regions. Without international trade, they may have little ability to benefit from comparative advantage, slicing up the value chain, or economies of scale. Moreover, smaller economies often have fewer competitive firms making goods within their economy, and thus firms have less pressure from other firms to provide the goods and prices that consumers want.

The economic gains from expanding international trade are measured in hundreds of billions of dollars, and the gains from international trade as a whole probably reach well into the trillions of dollars. The potential for gains from trade may be especially high among the smaller and lower-income countries of the world.

Link It Up

Visit this website for a list of some benefits of trade.

From Interpersonal to International Trade

Most people find it easy to believe that they, personally, would not be better off if they tried to grow and process all of their own food, to make all of their own clothes, to build their own cars and houses from scratch, and so on. Instead, we all benefit from living in economies where people and firms can specialize and trade with each other.

The benefits of trade do not stop at national boundaries, either. Earlier we explained that the division of labor could increase output for three reasons: (1) workers with different characteristics can specialize in the types of production where they have a comparative advantage; (2) firms and workers who specialize in a certain product become more productive with learning and practice; and (3) economies of scale. These three reasons apply from the individual and community level right up to the international level. If it makes sense to you that interpersonal, intercommunity, and interstate trade offer economic gains, it should make sense that international trade offers gains, too.

International trade currently involves about $20 trillion worth of goods and services moving around the globe. Any economic force of that size, even if it confers overall benefits, is certain to cause disruption and controversy. This module has only made the case that trade brings economic benefits. Other modules discuss, in detail, the public policy arguments over whether to restrict international trade.

It is Apple's (Global) iPhone

Apple Corporation uses a global platform to produce the iPhone. Now that you understand the concept of comparative advantage, you can see why the engineering and design of the iPhone is done in the United States. The United States has built up a comparative advantage over the years in designing and marketing products, and sacrifices fewer resources to design high-tech devices relative to other countries. China has a comparative advantage in assembling the phone due to its large skilled labor force. Korea has a comparative advantage in producing components. Korea focuses its production by increasing its scale, learning better ways to produce screens and computer chips, and uses innovation to lower average costs of production. Apple, in turn, benefits because it can purchase these quality products at lower prices. Put the global assembly line together and you have the device with which we are all so familiar.

Self Check: Comparative and Absolute Advantage

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the five Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/692

269. Outcome: Impact of International Trade

What you'll learn to do: explain how a nation's workers and consumers are affected by impact of international trade

In this section, you will examine international trade using the standard supply and demand framework.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Demand and Supply Analysis of International Trade
- Self Check: Impact of International Trade

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

270. Reading: Demand and Supply Analysis of International Trade

Demand and Supply Analysis of International Trade

The theories of comparative advantage and absolute advantage show us that there are overall gains from trade. Trade does have distributional impacts however. These distributional impacts are easier to see if one was to represent free trade in a standard demand and supply framework.

Consider two countries, Brazil and the United States, who produce sugar. Each country has a domestic supply and demand for sugar, as detailed in Table 1 and illustrated in Figure 2. In Brazil, without trade, the equilibrium price of sugar is 12 cents per pound and the equilibrium output is 30 tons. When there is no trade in the United States, the equilibrium price of sugar is 24 cents per pound and the equilibrium quantity is 80 tons. These equilibrium points are labeled with the point E. Notice that in this set-up, Brazil is the low-cost provider of sugar and has the cost-advantage.

Table 1. The Sugar Trade between Brazil and the United States

Price	Brazil: Quantity Supplied (tons)	Brazil: Quantity Demanded (tons)	U.S.: Quantity Supplied (tons)	U.S.: Quantity Demanded (tons)
8 cents	20	35	60	100
12 cents	30	30	66	93
14 cents	35	28	69	90
16 cents	40	25	72	87
20 cents	45	21	76	83
24 cents	50	18	80	80
28 cents	55	15	82	78

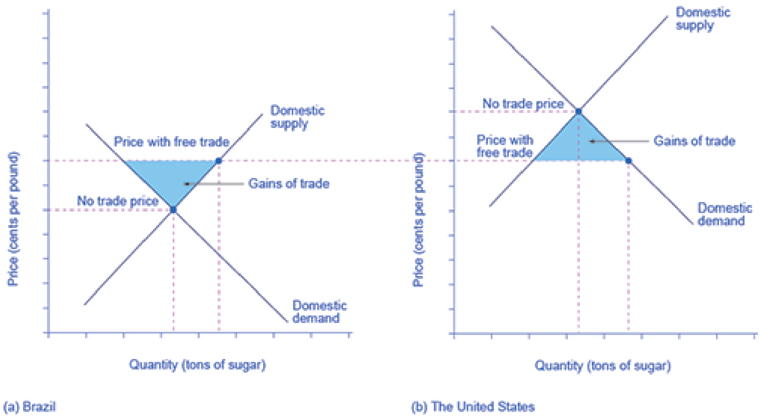


Figure 2. Free trade results in gains from trade. Total surplus increases in both countries. However, there are clear income distribution effects.

If international trade between Brazil and the United States now becomes possible, profit-seeking firms will spot an opportunity: buy sugar cheaply in Brazil, and sell it at a higher price in the United States. As sugar is shipped from Brazil to the United States, the quantity of sugar produced in Brazil will be greater than Brazilian consumption (with the extra production being exported), and the amount produced in the United States will be less than the amount of U.S. consumption (with the extra consumption being imported). Exports to the United States will reduce the supply of sugar in Brazil, raising its price. Imports into the United States will increase the supply of sugar, lowering its price. When the price of sugar is the same in both countries, there is no incentive to trade further. As Figure 34.2 shows, the equilibrium with trade occurs at a price of 16 cents per pound. At that price, the sugar farmers of Brazil supply a quantity of 40 tons, while the consumers of Brazil buy only 25 tons.

The extra 15 tons of sugar production, shown by the horizontal gap between the demand curve and the supply curve in Brazil, is exported to the United States. In the United States, at a price of 16 cents, the farmers produce a quantity of 72 tons and consumers demand a quantity of 87 tons. The excess demand of 15 tons by American consumers, shown by the horizontal gap between demand and domestic supply at the price of 16 cents, is supplied by imported sugar.

Free trade typically results in income distribution effects, but the key is to recognize the overall gains from trade, as shown in Figure 34.3. Building on the concepts outlined in Demand and Supply and Demand, Supply, and Efficiency (http://cnx.org/content/m48832/latest/) in terms of consumer and producer surplus, Figure 34.3 (a) shows that producers in Brazil gain by selling more sugar at a higher price, while Figure 34.3 (b) shows consumers in the United States benefit from the lower price and greater availability of sugar. Consumers in Brazil are worse off (compare their no- trade consumer surplus with the free-trade consumer surplus) and U.S. producers of sugar are worse off. There are gains from trade—an increase in social surplus in each country. That is, both the United States and Brazil are better off than they would be without trade.

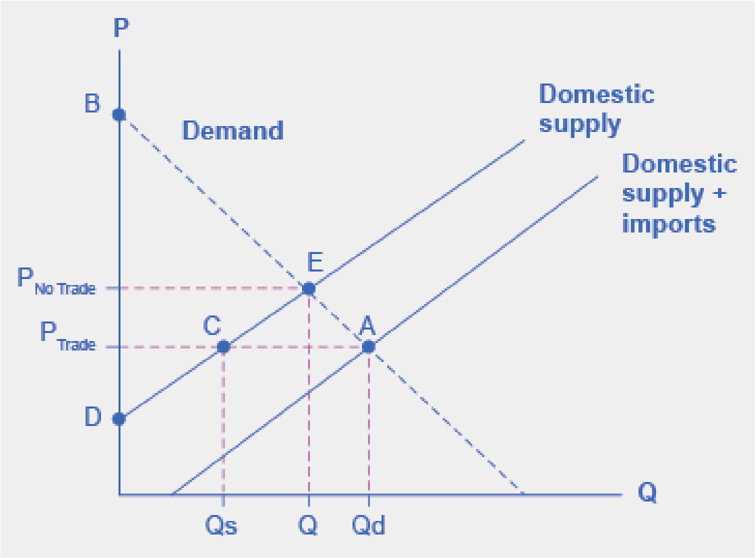


Figure 3. When there is free trade, the equilibrium is at point A. When there is no trade, the equilibrium is at point E.

The fact that there are distributional consequences to trade is exactly the reason why workers and business lobby government for trade restrictions and protectionist regulations.

Self Check: Impact of International Trade

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether

to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/693

271. Outcome: Impact of Government Regulations

What you'll learn to do: understand the way government regulations (like tariffs, quotas and non-tariff barriers) affect businesses, consumers and workers in the economy

In this section, you will analyze the ripple effects of government regulations on the economy.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Restrictions on International Trade
- Video: Why Do Countries Restrict Trade?
- Reading: Justifications for Trade Restriction
- Video: Types of Trade Restrictions
- Case in Point: Outsourcing, Insourcing, and Employment
- Reading: The Tradeoffs of Trade Policy
- Self Check: Impacts of Government Regulations

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

272. Reading: Restrictions on International Trade

Restrictions on International Trade

In spite of the strong theoretical case that can be made for free international trade, every country in the world has erected at least some barriers to trade. Trade restrictions are typically undertaken in an effort to protect companies and workers in the home economy from competition by foreign firms. A protectionist policy is one in which a country restricts the importation of goods and services produced in foreign countries. The slowdown in the U.S. economy late in 2007 and in 2008 has produced a new round of protectionist sentiment—one that became a factor in the 2008 U.S. presidential campaign.

The United States, for example, uses protectionist policies to limit the quantity of foreign-produced sugar coming into the United States. The effect of this policy is to reduce the supply of sugar in the U.S. market and increase the price of sugar in the United States. The 2008 U.S. Farm Bill sweetened things for sugar growers even more. It raised the price they are guaranteed to receive and limited imports of foreign sugar so that American growers will always have at least 85% of the domestic market. The bill for the first time set an income limit—only growers whose incomes fall below $1.5 million per year (for couples) or $750,000 for individuals will receive direct subsidies ([1]

The U.S. price of sugar is almost triple the world price of sugar,

1. "Who Wants to Be a Millionaire?" *The Wall Street Journal*, May 14, 2008, p. A20).

thus reducing the quantity consumed in the United States. The program benefits growers of sugar beets and sugar cane at the expense of consumers.

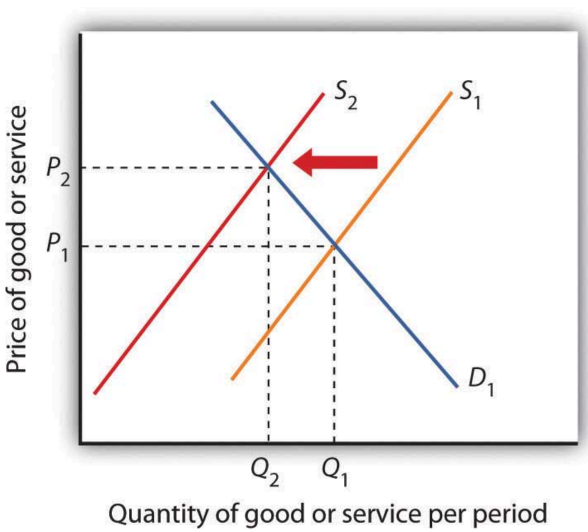


Figure 17.7. *The Impact of Protectionist Policies. Protectionist policies reduce the quantities of foreign goods and services supplied to the country that imposes the restriction. As a result, such policies shift the supply curve to the left for the good or service whose imports are restricted. In the case shown, the supply curve shifts to S2, the equilibrium price rises to P2, and the equilibrium quantity falls to Q2.*

In general, protectionist policies imposed for a particular good always reduce its supply, raise its price, and reduce the equilibrium quantity, as shown in Figure 17.7 “The Impact of Protectionist Policies”. Protection often takes the form of an import tax or a limit on the amount that can be imported, but it can also come in the form of voluntary export restrictions and other barriers.

Tariffs

A tariff is a tax on imported goods and services. The average tariff on dutiable imports in the United States (that is, those imports on which a tariff is imposed) is about 4%. Some imports have much higher tariffs. For example, the U.S. tariff on imported frozen orange juice is 35 cents per gallon (which amounts to about 40% of value). The tariff on imported canned tuna is 35%, and the tariff on imported shoes ranges between 2% and 48%.

A tariff raises the cost of selling imported goods. It thus shifts the supply curve for goods to the left, as in Figure 17.7 “The Impact of Protectionist Policies.” The price of the protected good rises and the quantity available to consumers falls.

Antidumping Proceedings

One of the most common protectionist measures now in use is the antidumping proceeding. A domestic firm, faced with competition by a foreign competitor, files charges with its government that the foreign firm is dumping, or charging an “unfair” price. Under rules spelled out in international negotiations that preceded approval of the World Trade Organization, an unfair price was defined as a price below production cost or below the price the foreign firm charges for the same good in its own country. While these definitions may seem straightforward enough, they have proven to be quite troublesome. The definition of “production cost” is a thoroughly arbitrary procedure. In defining cost, the government agency invariably includes a specification of a “normal” profit. That normal profit can be absurdly high. The United States Department of Justice, which is the U.S. agency in charge of determining whether a foreign firm has charged an unfair price, has sometimes defined normal profit rates as exceeding production cost by well over 50%, a rate far higher than exists in most U.S. industry.

The practice of a foreign firm charging a price in the United States that is below the price it charges in its home country is common. The U.S. market may be more competitive, or the foreign firm may simply be trying to make its product attractive to U.S. buyers that are not yet accustomed to its product. In any event, such price discrimination behavior is not unusual and is not necessarily “unfair.”

In the United States, once the Department of Justice has determined that a foreign firm is guilty of charging an unfair price, the U.S. International Trade Commission must determine that the foreign firm has done material harm to the U.S. firm. If a U.S. firm has suffered a reduction in sales and thus in employment it will typically be found to have suffered material harm, and punitive duties will be imposed.

Quotas

A quota is a direct restriction on the total quantity of a good or service that may be imported during a specified period. Quotas restrict total supply and therefore increase the domestic price of the good or service on which they are imposed. Quotas generally specify that an exporting country’s share of a domestic market may not exceed a certain limit.

In some cases, quotas are set to raise the domestic price to a particular level. Congress requires the Department of Agriculture, for example, to impose quotas on imported sugar to keep the wholesale price in the United States above 22 cents per pound. The world price is typically less than 10 cents per pound.

A quota restricting the quantity of a particular good imported into an economy shifts the supply curve to the left, as in Figure 17.7 "The Impact of Protectionist Policies." It raises price and reduces quantity.

An important distinction between quotas and tariffs is that quotas do not increase costs to foreign producers; tariffs do. In the short run, a tariff will reduce the profits of foreign exporters of a good or service. A quota, however, raises price but not costs of production and thus may increase profits. Because the quota imposes a limit on quantity, any profits it creates in other countries will not induce the entry of new firms that ordinarily eliminates profits in perfect competition. By definition, entry of new foreign firms to earn the profits available in the United States is blocked by the quota.

Voluntary Export Restrictions

Voluntary export restrictions are a form of trade barrier by which foreign firms agree to limit the quantity of goods exported to a particular country. They became prominent in the United States in the 1980s, when the U.S. government persuaded foreign exporters of automobiles and steel to agree to limit their exports to the United States.

Although such restrictions are called voluntary, they typically are agreed to only after pressure is applied by the country whose industries they protect. The United States, for example, has succeeded in pressuring many other countries to accept quotas limiting their exports of goods ranging from sweaters to steel.

A voluntary export restriction works precisely like an ordinary quota. It raises prices for the domestic product and reduces the quantity consumed of the good or service affected by the quota. It can also increase the profits of the firms that agree to the quota because it raises the price they receive for their products.

Other Barriers

In addition to tariffs and quotas, measures such as safety standards, labeling requirements, pollution controls, and quality restrictions all may have the effect of restricting imports.

Many restrictions aimed at protecting consumers in the domestic market create barriers as a purely unintended, and probably desirable, side effect. For example, limitations on insecticide levels in foods are often more stringent in the United States than in other countries. These standards tend to discourage the import of foreign goods, but their primary purpose appears to be to protect consumers from harmful chemicals, not to restrict trade. But other nontariff barriers seem to serve no purpose other than to keep foreign goods out. Tomatoes produced in Mexico, for example, compete with those produced in the United States. But Mexican tomatoes tend to be smaller than U.S. tomatoes. The United States once imposed size restrictions to "protect" U.S. consumers from small tomatoes. The result was a highly effective trade barrier that protected U.S. producers and raised U.S. tomato prices. Those restrictions were abolished under terms of the North American Free Trade Agreement, which has led to a large increase in U.S. imports of Mexican tomatoes and a reduction in U.S. tomato production.

Trade Restrictions Video

Why do countries restrict trade? Watch this video to learn about the major arguments in favor of trade restrictions, including: protecting domestic jobs, leveling the playing field, providing government revenue, supporting national defense, protecting national interests, protecting infant industries, and promoting exports.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=304#oembed-1

Types of Trade Restrictions Video

This video details the ways a country might restrict trade and explains the impact of these restrictions.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=304#oembed-2

273. Reading: Justifications for Trade Restriction

Justifications for Trade Restriction: An Evaluation

The conceptual justification for free trade is one of the oldest arguments in economics; there is no disputing the logic of the argument that free trade increases global production, worldwide consumption, and international efficiency. But critics stress that the argument is a theoretical one. In the real world, they say, there are several arguments that can be made to justify protectionist measures.

Infant Industries

One argument for trade barriers is that they serve as a kind of buffer to protect fledgling domestic industries. Initially, firms in a new industry may be too small to achieve significant economies of scale and could be clobbered by established firms in other countries. A new domestic industry with potential economies of scale is called an infant industry.

Consider the situation in which firms in a country are attempting to enter a new industry in which many large firms already exist in the international arena. The foreign firms have taken advantage of economies of scale and have therefore achieved relatively low levels of production costs. New firms, facing low levels of output and higher average costs, may find it difficult to compete. The infant industry argument suggests that by offering protection during an industry's formative years, a tariff or quota may allow the new industry to develop and prosper.

The infant industry argument played a major role in tariff policy in the early years of U.S. development. The high tariffs of the early nineteenth century were typically justified as being necessary to allow U.S. firms to gain a competitive foothold in the world economy. As domestic industries became established, tariff rates fell. Subsequent increases in tariffs were a response in part to internal crises: the Civil War and the Great Depression. Tariff rates have fallen dramatically since 1930.

Critics of the infant industry argument say that once protection is in place, it may be very difficult to remove. Inefficient firms, they contend, may be able to survive for long periods under the umbrella of infant industry protection.

Strategic Trade Policy

A new version of the infant industry argument has been used in the past few years as technological developments have spawned whole new industries and transformed existing ones. The new version of the infant industry argument assumes an imperfectly competitive market.

Suppose technological change has given rise to a new industry. Given the economies of scale in this industry, only a few firms are likely to dominate it worldwide—it will likely emerge as an oligopoly. The firms that dominate the industry are likely to earn economic profits that will persist. Furthermore, because there will be only a few firms, they will be located in only a few countries. Their governments could conceivably impose taxes on these firms' profits that would enhance economic well-being within the country. The potential for such gains may justify government efforts to assist firms seeking to acquire a dominant position in the new industry.

Government aid could take the form of protectionist trade policies aimed at allowing these firms to expand in the face of foreign competition, assistance with research and development efforts, programs to provide workers with special skills needed by the industry, or subsidies in the form of direct payments or special tax treatment. Any such policy aimed at promoting the development of key industries that may increase a country's domestic well-being through trade with the rest of the world is known as a strategic trade policy.

Although strategic trade policy suggests a conceptually positive role for government in international trade, proponents of the approach note that it has dangers. Firms might use the strategic trade argument even if their development were unlikely to offer the gains specified in the theory. The successful application of the approach requires that the government correctly identify industries in which a country can, in fact, gain dominance—something that may not be possible. Various European governments provided subsidies to firms that were involved in the production of Airbus, which is now a major competitor in the airplane industry. On the other hand, Britain and France subsidized the development of the supersonic plane called the Concorde. After only a few Concordes had been produced, it became obvious that the aircraft was a financially losing proposition and production was halted. The airline has now gone out of business.

Finally, those firms whose success strategic trade policy promotes might have sufficient political clout to block the taxes that would redistribute the gains of the policies to the population in general. Thus, the promise of strategic trade policy is unlikely to be fulfilled.

National Security And the National Interest Argument

Some argue that a nation should not depend too heavily on other countries for supplies of certain key products, such as oil, or for special materials or technologies that might have national security applications. On closer consideration, this argument for protectionism proves rather weak.

As an example, in the United States, oil provides about 40% of all the energy and about 40% of the oil used in the United States economy is imported. Several times in the last few decades, when disruptions in the Middle East have shifted the supply curve of oil back to the left and sharply raised the price, the effects have been felt across the United States economy. This is not, however, a very convincing argument for restricting imports of oil. If the United States needs to be protected from a possible cutoff of foreign oil, then a more reasonable strategy would be to import 100% of the petroleum supply now, and save U.S. domestic oil resources for when or if the foreign supply is cut off. It might also be useful to import extra oil and put it into a stockpile for use in an emergency, as the United States government did by starting a Strategic Petroleum Reserve in 1977. Moreover, it may be necessary to discourage people from using oil, and to start a high-powered program to seek out alternatives to oil. A straightforward way to do this would be to raise taxes on oil. What's more, it makes no sense to argue that because oil is highly important to the United States economy, then the United States should shut out oil imports and use up its domestic supplies of oil more quickly.

Whether or not to limit certain kinds of imports of key technologies or materials that might be important to national security and weapons systems is a slightly different issue. If weapons' builders are not confident that they can continue to obtain a key product in wartime, they might decide to avoid designing weapons that use this key product, or they can go ahead and design the weapons and stockpile enough of the key high-tech components or materials to last through an armed conflict. Indeed, there is a U.S. Defense National Stockpile Center that has built up reserves of many materials, from aluminum oxides, antimony, and bauxite to tungsten, vegetable tannin extracts, and zinc (although many of these stockpiles have been reduced and sold in recent years).

One final reason why economists often treat the *national interest argument* skeptically is that almost any product can be touted by lobbyists and politicians as vital to national security. In 1954, the United States became worried that it was importing half of the wool required for military uniforms, so it declared wool and mohair to be "strategic materials" and began to give subsidies to wool and mohair farmers. Although wool was removed from the official list of "strategic" materials in 1960, the subsidies for mohair continued for almost 40 years until they were repealed in 1993, and then were reinstated in 2002. All too often, the national interest argument has become an excuse for handing out the indirect subsidy of protectionism to certain industries or companies. After all, decisions about what constitutes a key strategic material are made by politicians, not nonpartisan analysts.

Job Protection

The desire to maintain existing jobs threatened by foreign competition is probably the single most important source of today's protectionist policies. Some industries that at one time had a comparative advantage are no longer among the world's lowest-cost producers; they struggle to stay afloat. Cost cutting leads to layoffs, and layoffs lead to demands for protection.

The model of international trade in perfect competition suggests that trade will threaten some industries. As countries specialize in activities in which they have a comparative advantage, sectors in which they do not have this advantage will shrink. Maintaining those sectors through trade barriers blocks a nation from enjoying the gains possible from free trade.

A further difficulty with the use of trade barriers to shore up employment in a particular sector is that it can be an enormously expensive strategy. Suppose enough of a foreign good is kept out of the United States to save one U.S. job. That shifts the supply curve slightly to the left, raising prices for U.S. consumers and reducing their consumer surplus. The loss to consumers is the cost per job saved. Estimates of the cost of saving *one* job in the steel industry through restrictions on steel imports, for example, go as high as $800,000 per year.

Cheap Foreign Labor and Outsourcing

One reason often given for the perceived need to protect American workers against free international trade is that workers must be protected against cheap foreign labor. This is an extension of the job protection argument in the previous section. From a theoretical point of view, of course, if foreign countries can produce a good at lower cost than we can, it is in our collective interest to obtain it from them. But workers counter by saying that the low wages of foreign workers means that foreign workers are exploited. To compete with foreign workers, American workers would have to submit themselves to similar exploitation. This objection, however, fails to recognize that differences in wage rates generally reflect differences in worker productivity.

Consider the following example: Suppose U.S. workers in the tool industry earn $20 per hour while Indonesian workers in the tool industry earn only $2 per hour. If we assume that the tool industry is competitive, then the wages in both countries are based on the marginal revenue product of the workers. The higher wage of U.S. workers must mean that they have a higher marginal product—they are more productive. The higher wage of U.S. workers need not mean that labor costs are higher in the United States than in Indonesia.

Further, we have seen that what matters for trade is comparative advantage, not comparative labor costs. When each nation specializes in goods and services in which it has a comparative advantage—measured in the amounts of other goods and services given up to produce them—then world production, and therefore world consumption, rises. By definition, each nation will have a comparative advantage in *something.*

A particularly controversial issue in industrialized economies is outsourcing, in which firms in a developed country transfer some of their activities abroad in order to take advantage of lower labor costs in other countries. Generally speaking, the practice of outsourcing tends to reduce costs for the firms that do it. These firms often expand production and increase domestic employment, as is discussed in the accompanying Case in Point essay.

Differences in Environmental Standards

Another justification for protectionist measures is that free trade is unfair if it pits domestic firms against foreign rivals who do not have to adhere to the same regulatory standards. In the debate over NAFTA, for example, critics warned that Mexican firms, facing relatively lax pollution control standards, would have an unfair advantage over U.S. firms if restraints on trade between the two countries were removed.

Economic theory suggests, however, that differences in pollution-control policies can be an important source of comparative advantage. In general, the demand for environmental quality is positively related to income. People in higher-income countries demand higher environmental quality than do people in lower-income countries. That means that pollution has a lower cost in poorer than in richer countries. If an industry generates a great deal of pollution, it may be more efficient to locate it in a poor country than in a rich country. In effect, a poor country's lower demand for environmental quality gives it a comparative advantage in production of goods that generate a great deal of pollution.

Provided the benefits of pollution exceed the costs in the poor country, with the costs computed based on the preferences and incomes of people in that country, it makes sense for more of the good to be produced in the poor country and less in the rich country. Such an allocation leaves people in both countries better off than they would be otherwise. Then, as freer trade leads to higher incomes in the poorer countries, people there will also demand improvements in environmental quality.

Do economists support *any* restriction on free international trade? Nearly all economists would say no. The gains from trade are so large, and the cost of restraining it so high, that it is hard to find any satisfactory reason to limit trade.

274. Case in Point: Outsourcing, Insourcing, and Employment

Outsourcing, Insourcing, and Employment

The phenomenon of outsourcing has become common as the Internet and other innovations in communication have made it easier for firms to transfer aspects of their production overseas. At the same time, countries such as India and China have invested heavily in education and have produced a sizable workforce of professional people capable of filling relatively high level positions for firms in more developed countries.

The very idea of outsourcing rankles politicians on the left and on the right. In the United States, there have been numerous congressional hearings on outsourcing and proposals to block firms that engage in the practice from getting government contracts.

By outsourcing, firms are able to reduce their production costs. As we have seen, a reduction in production costs translates into increased output and falling prices. From a consumer's point of view, then, outsourcing should be a very good thing. The worry many commentators express, however, is that outsourcing will decimate employment in the United States, particularly among high-level professionals. Matthew J. Slaughter, an economist at Dartmouth University, examined employment trends from 1991 to 2001 among multinational U.S. firms that had outsourced jobs. Those firms outsourced 2.8 million jobs during the period.

Were the 2.8 million jobs simply lost? Mr. Slaughter points out that there are three reasons to expect that the firms that reduced production costs by outsourcing would actually increase their domestic employment. First, by lowering cost, firms are likely to expand the quantity they produce. The foreign workers who were hired, who Mr. Slaughter refers to as "affiliate workers," appeared to be complements to American workers rather than substitutes. If they are complements rather than substitutes, then outsourcing could lead to increased employment in the country that does the outsourcing.

A second reason outsourcing could increase employment is that by lowering production cost, firms that increase the *scale* of their operations through outsourcing need more domestic workers to sell the increased output, to coordinate its distribution, and to develop the infrastructure to handle all those goods.

Finally, firms that engage in outsourcing are also likely to increase the *scope* of their operations. They will need to hire additional people to explore other product development, to engage in research, and to seek out new markets for the firm's output.

Thus, Mr. Slaughter argues that outsourcing may lead to increased employment because domestic workers are complements to foreign workers, because outsourcing expands the scale of a firm's operations, and because it expands the scope of operations. What did the evidence show? Remember the 2.8 million jobs that multinational firms based in the United States outsourced between 1991 and 2001? Employment at those same U.S. firms *increased* by 5.5 million jobs during the period. Thus, with the phenomena of complementarity, increases in scale, and increases of scope, each job outsourced led to almost two *additional* jobs in the United States.

The experience of two quite dissimilar firms illustrates the phenomenon. Walmart began expanding its operations internationally in about 1990. Today, it manages its global operations from its headquarters in Bentonville, Arkansas where it employs 15,000 people. Roughly 1,500 of these people coordinate the flow of goods among Walmart's stores throughout the world. Those 1,500 jobs would not exist were it not for globalization. Xilinx, the high technology research and development firm, generates sales of about $1.5 billion per year. Sixty-five percent of its sales are generated outside the United States. But 80% of its employees are in the United States.

Outsourcing, then, generates jobs. It does not destroy them. Mr. Slaughter concludes: “Instead of lamenting ongoing foreign expansion of U.S. multinationals, if history is our guide then we should be encouraging it.”

Mr. Slaughter and co-researcher Robert Kimmitt make a similar case for insourcing, production, and jobs generated by multinationals based outside the United States that build plants inside the United States. In 2007, these companies employed almost 2 million Americans with average compensation of nearly $80,000. Contrary to popular belief, the unionization rates of the United States affiliates of these companies were about 50% higher than for the rest of the U.S. private sector.

Think every country is pro-trade? How about the U.S.? The following reading might surprise you.

HOW DOES THE UNITED STATES REALLY FEEL ABOUT EXPANDING TRADE?

How do people around the world feel about expanding trade between nations? In summer 2007, the Pew Foundation surveyed 45,000 people in 47 countries. One of the questions asked about opinions on growing trade ties between countries. Table 20.3 shows the percentages who answered either “very good” or “somewhat good” for some of countries surveyed.

For those who think of the United States as the world’s leading supporter of expanding trade, the survey results may be perplexing. When adding up the shares of those who say that growing trade ties between countries is “very good” or “somewhat good,” Americans had the least favorable attitude toward increasing globalization, while the Chinese and South Africans ranked highest. In fact, among the 47 countries surveyed, the United States ranked by far the lowest on this measure, followed by Egypt, Italy, and Argentina.

Country	Very Good	Somewhat Good	Total
China	38%	53%	91%
South Africa	42%	43%	87%
South Korea	24%	62%	86%
Germany	30%	55%	85%
Canada	29%	53%	82%
United Kingdom	28%	50%	78%
Mexico	22%	55%	77%
Brazil	13%	59%	72%
Japan	17%	55%	72%
United States	14%	45%	59%

Table 20.3. The Status of Growing Trade Ties between Countries (Source: http://www.pewglobal.org/files/pdf/258.pdf)

275. Reading: The Tradeoffs of Trade Policy

The Tradeoffs of Trade Policy

Economists readily acknowledge that international trade is not all sunshine, roses, and happy endings. Over time, the average person gains from international trade, both as a worker who has greater productivity and higher wages because of the benefits of specialization and comparative advantage, and as a consumer who can benefit from shopping all over the world for a greater variety of quality products at attractive prices. The "average person," however, is hypothetical, not real—representing a mix of those who have done very well, those who have done all right, and those who have done poorly. It is a legitimate concern of public policy to focus not just on the average or on the success stories, but also on those have not been so fortunate. Workers in other countries, the environment, and prospects for new industries and materials that might be of key importance to the national economy are also all legitimate issues.

The common belief among economists is that it is better to embrace the gains from trade, and then deal with the costs and tradeoffs with other policy tools, than it is to cut off trade to avoid the costs and tradeoffs.

To gain a better intuitive understanding for this argument, consider a hypothetical American company called Technotron. Technotron invents a new scientific technology that allows the firm to increase the output and quality of its goods with a smaller number of workers at a lower cost. As a result of this technology, other U.S. firms in this industry will lose money and will also have to lay off workers—and some of the competing firms will even go bankrupt. Should the United States government protect the existing firms and their employees by making it illegal for Technotron to use its new technology? Most people who live in market-oriented economies would oppose trying to block better products that lower the cost of services. Certainly, there is a case for society providing temporary support and assistance for those who find themselves without work. Many would argue for government support of programs that encourage retraining and acquiring additional skills. Government might also support research and development efforts, so that other firms may find ways of outdoing Technotron. Blocking the new technology altogether, however, seems like a mistake. After all, few people would advocate giving up electricity because it caused so much disruption to the kerosene and candle business. Few would suggest holding back on improvements in medical technology because they might cause companies selling leeches and snake oil to lose money. In short, most people view disruptions due to technological change as a necessary cost that is worth bearing.

Now, imagine that Technotron's new "technology" is as simple as this: the company imports what it sells from another country. In other words, think of foreign trade as a type of innovative technology. The objective situation is now exactly the same as before. Because of Technotron's new technology—which in this case is importing goods from another county—other firms in this industry will lose money and lay off workers. Just as it would have been inappropriate and ultimately foolish to respond to the disruptions of new scientific technology by trying to shut it down, it would be inappropriate and ultimately foolish to respond to the disruptions of *international trade* by trying to restrict trade.

Some workers and firms will suffer because of international trade. In a living, breathing market-oriented economy, some workers and firms will always be experiencing disruptions, for a wide variety of reasons. Corporate management can be better or worse. Workers for a certain firm can be more productive or less. Tough domestic competitors can create just as much disruption as tough foreign competitors. Sometimes a new product is a hit with consumers; sometimes it is a flop. Sometimes a company is blessed by a run of good luck or stricken with a run of bad luck. For some firms, international trade will offer great opportunities for expanding productivity and jobs; for other firms, trade will impose stress and pain. The disruption caused by international trade is not fundamentally different from all the other disruptions caused by the other workings of a market economy.

In other words, the economic analysis of free trade does not rely on a belief that foreign trade is not disruptive or does not pose tradeoffs; indeed, the story of Technotron begins with a particular *disruptive market change*—a new technology—that causes real tradeoffs. In thinking about the disruptions of foreign trade, or any of the other possible costs and tradeoffs of foreign trade discussed in this module, the best public policy solutions typically do not involve protectionism, but instead involve finding ways for public policy to address the particular issues, while still allowing the benefits of international trade to occur.

WHAT'S THE DOWNSIDE OF PROTECTION?

Flat-panel displays, the displays for laptop computers, tablets, and flat screen televisions, are an example of such an enduring principle. In the early 1990s, the vast majority of flat-panel displays used in U.S.-manufactured laptops were imported, primarily from Japan. The small but politically powerful U.S. flat-panel-display industry filed a dumping complaint with the Commerce Department. They argued that Japanese firms were selling displays at "less than fair value," which made it difficult for U.S. firms to compete. After a preliminary determination by the Commerce Department that the Japanese firms were dumping, the U.S. International Trade Commission imposed a 63% dumping margin (or tax) on the import of flat-panel displays. Was this a successful exercise of U.S. trade policy?

Flat-panel displays make up a significant portion of the cost of producing laptop computers—as much as 50%. Therefore, the antidumping tax would substantially increase the cost, and thus the price, of U.S.-manufactured laptops. As a result of the ITC's decision, Apple moved its domestic manufacturing plant for Macintosh computers to Ireland (where it had an existing plant). Toshiba shut down its U.S. manufacturing plant for laptops. And IBM cancelled plans to open a laptop manufacturing plant in North Carolina, instead deciding to expand production at its plant in Japan. In this case, rather than having the desired effect of protecting U.S. interests and giving domestic manufacturing an advantage over items manufactured elsewhere, it had the unintended effect of driving the manufacturing completely out of the country. Many people lost their jobs and most flat-panel display production now occurs in countries other than the United States.

Self Check: Impact of Government Regulations

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the three Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/694

276. Outcome: Trade Policy and Agreements

What you'll learn to do: differentiate between alternative international trade regimes and how they impact global trade

In this outcome, you will examine the World Trade Organization and regional trade associations to see how they impact global trade relations.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: How Trade Policy is Enacted
- Video: Trade Agreements
- Video: South Korea Free Trade Agreement
- Self Check: Trade Policy and Agreements

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

277. Reading: How Trade Policy Is Enacted

How Trade Policy Is Enacted: Globally, Regionally, and Nationally

These public policy arguments about how nations should react to globalization and trade are fought out at several levels: at the global level through the World Trade Organization and through regional trade agreements between pairs or groups of countries.

The World Trade Organization

The *World Trade Organization* (WTO) was officially born in 1995, but its history is much longer. In the years after the Great Depression and World War II, there was a worldwide push to build institutions that would tie the nations of the world together. The United Nations officially came into existence in 1945. The World Bank, which assists the poorest people in the world, and the International Monetary Fund, which addresses issues raised by international financial transactions, were both created in 1946. The third planned organization was to be an International Trade Organization, which would manage international trade. The United Nations was unable to agree to this. Instead, the *General Agreement on Tariffs and Trade* (GATT), was established in 1947 to provide a forum in which nations could come together to negotiate reductions in tariffs and other barriers to trade. In 1995, the GATT was transformed into the WTO.

The GATT process was to negotiate an agreement to reduce barriers to trade, sign that agreement, pause for a while, and then start negotiating the next agreement. The rounds of talks in the GATT, and now the WTO, are shown in Table 20.1. Notice that the early rounds of GATT talks took a relatively short time, included a small number of countries, and focused almost entirely on reducing tariffs. Since the 1970s, however, rounds of trade talks have taken years, included a large number of countries, and an ever-broadening range of issues.

Table 20.1 The Negotiating Rounds of GATT and the World Trade Organization

Year	Place or Name of Round	Main Subjects	Number of Countries Involved
1947	Geneva	Tariff reduction	23
1949	Annecy	Tariff reduction	13
1951	Torquay	Tariff reduction	38
1956	Geneva	Tariff reduction	26
1960–61	Dillon round	Tariff reduction	26
1964–67	Kennedy round	Tariffs, anti-dumping measures	62
1973–79	Tokyo round	Tariffs, nontariff barriers	102
1986–94	Uruguay round	Tariffs, nontariff barriers, services, intellectual property, dispute settlement, textiles, agriculture, creation of WTO	123
2001–	Doha round	Agriculture, services, intellectual property, competition, investment, environment, dispute settlement	147

The sluggish pace of GATT negotiations led to an old joke that GATT really stood for Gentleman's Agreement to Talk and Talk. The slow pace of international trade talks, however, is understandable, even sensible. Having dozens of nations agree to any treaty is a lengthy process. GATT often set up separate trading rules for certain industries, like agriculture, and separate trading rules for certain countries, like the low-income countries. There were rules, exceptions to rules, opportunities to opt out of rules, and precise wording to be fought over in every case. Like the GATT before it, the WTO is not a world government, with power to impose its decisions on others. The total staff of the WTO in 2013 is 629 people and its annual budget (as of 2012) is $196 million, which makes it smaller in size than many large universities.

Regional Trading Agreements

There are different types of economic integration across the globe, ranging from *free trade agreements*, in which participants allow each other's imports without tariffs or quotas, to *common markets*, in which participants have a common external trade policy as well as free trade within the group, to full *economic unions*, in which, in addition to a common market, monetary and fiscal policies are coordinated. Many nations belong both to the World Trade Organization and to regional trading agreements.

The best known of these regional trading agreements is the *European Union*. In the years after World War II, leaders of several European nations reasoned that if they could tie their economies together more closely, they might be more likely to avoid another devastating war. Their efforts began with a free trade association, evolved into a common market, and then transformed into what is now a full economic union, known as the European Union. The EU, as it is often called, has a number of goals. For example, in the early 2000s it introduced a common currency for Europe, the euro, and phased out most of the former national forms of money like the German mark and the French franc, though a few have retained their own currency. Another key element of the union is to eliminate barriers to the mobility of goods, labor, and capital across Europe.

For the United States, perhaps the best-known regional trading agreement is the *North American Free Trade Agreement* (NAFTA).

The United States also participates in some less-prominent regional trading agreements, like the Caribbean Basin Initiative, which offers reduced tariffs for imports from these countries, and a free trade agreement with Israel.

The world has seen a flood of regional trading agreements in recent years. About 100 such agreements are now in place. A few of the more prominent ones are listed in Table 20.2. Some are just agreements to continue talking; others set specific goals for reducing tariffs, import quotas, and nontariff barriers. One economist described the current trade treaties as a “spaghetti bowl,” which is what a map with lines connecting all the countries with trade treaties looks like.

There is concern among economists who favor free trade that some of these regional agreements may promise free trade, but actually act as a way for the countries within the regional agreement to try to limit trade from anywhere else. In some cases, the regional trade agreements may even conflict with the broader agreements of the World Trade Organization.

Table 20.2. Some Regional Trade Agreements

Trade Agreements	Participating Countries
Asia Pacific Economic Cooperation (APEC)	Australia, Brunei, Canada, Chile, People's Republic of China, Hong Kong, China, Indonesia, Japan, Republic of Korea, Malaysia, Mexico, New Zealand, Papua New Guinea, Peru, Philippines, Russia, Singapore, Chinese Taipei, Thailand, United States, Vietnam
European Union (EU)	Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom
North America Free Trade Agreement (NAFTA)	Canada, Mexico, United States
Latin American Integration Association (LAIA)	Argentina, Bolivia, Brazil, Chile, Columbia, Ecuador, Mexico, Paraguay, Peru, Uruguay, Venezuela
Association of Southeast Asian Nations (ASEAN)	Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Vietnam
Southern African Development Community (SADC)	Angola, Botswana, Congo, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Zambia, Zimbabwe

Yet another dimension of trade policy, along with international and regional trade agreements, happens at the national level. The United States, for example, imposes import quotas on sugar, because of a fear that such imports would drive down the price of sugar and thus injure domestic sugar producers. One of the jobs of the United States Department of Commerce is to determine if imports from other countries are being dumped. The United States International Trade Commission—a government agency—determines whether domestic industries have been substantially injured by the dumping, and if so, the president can impose tariffs that are intended to offset the unfairly low price.

In the arena of trade policy, the battle often seems to be between national laws that increase protectionism and international agreements that try to reduce protectionism, like the WTO. Why would a country pass laws or negotiate agreements to shut out certain foreign products, like sugar or textiles, while simultaneously negotiating to reduce trade barriers in general? One plausible answer is that international trade agreements offer a method for countries to restrain their own special interests. A member of Congress can say to an industry lobbying for tariffs or quotas on imports: "Sure would like to help you, but that pesky WTO agreement just won't let me."

Link It Up

If consumers are the biggest losers from trade, why do they not fight back? The quick answer is because it is easier to organize a small group of people around a narrow interest versus a large group that has diffuse interests. This is a question about trade policy theory. Visit this website and read the article by Jonathan Rauch.

Long-Term Trends in Barriers to Trade

In newspaper headlines, trade policy appears mostly as disputes and acrimony. Countries are almost constantly threatening to challenge the "unfair" trading practices of other nations. Cases are brought to the dispute settlement procedures of the WTO, the European Union, NAFTA, and other regional trading agreements. Politicians in national legislatures, goaded on by lobbyists, often threaten to pass bills that will "establish a fair playing field" or "prevent unfair trade"—although most such bills seek to accomplish these high-sounding goals by placing more restrictions on trade. Protesters in the streets may object to specific trade rules or to the entire practice of international trade.

Through all the controversy, the general trend in the last 60 years is clearly toward lower barriers to trade. The average level of tariffs on imported products charged by industrialized countries was 40% in 1946. By 1990, after decades of GATT negotiations, it was down to less than 5%. Indeed, one of the reasons that GATT negotiations shifted from focusing on tariff reduction in the early rounds to a broader agenda was that tariffs had been reduced so dramatically there was not much more to do in that area. U.S. tariffs have followed this general pattern: After rising sharply during the Great Depression, tariffs dropped off to less than 2% by the end of the century. Although measures of import quotas and nontariff barriers are less exact than those for tariffs, they generally appear to be at lower levels, too.

Thus, the last half-century has seen both a dramatic reduction in government-created barriers to trade, such as tariffs, import quotas, and nontariff barriers, and also a number of technological developments that have made international trade easier, like advances in transportation, communication, and information management. The result has been the powerful surge of international trade.

278. Videos: Trade Agreements

Trade Agreements

Watch these videos about trade agreements and bloc to better understand trade relationships between nations.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=310#oembed-1

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=310#oembed-2

South Korea Free Trade Agreement

Watch this video to learn about U.S.-Korea Free Trade Agreement signed in 2012.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=310#oembed-3

279. Self Check: Trade Policy and Agreements

Check Your Understanding

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the Reading in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/695

280. Outcome: Exchange Rates and International Finance

What you'll learn to do: define currency exchange rates and explain how they influence trade balances

In this section, you will take a look at the foreign exchange market and analyze how different currencies necessitate some degree of consistency for purposes of international trade.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Exchange Rates and International Capital Flows
- Reading: The Foreign Exchange Market
- Reading: Strengthening and Weakening Currency
- Reading: Demand and Supply Shifts in Foreign Exchange Markets
- Self Check: Exchange Rates and International Finance

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

281. Reading: Exchange Rates and International Capital Flows

Figure 15.1. *Trade Around the World. Is a trade deficit between the United States and the European Union good or bad for the U.S. economy? (Credit: modification of work by Milad Mosapoor/ Wikimedia Commons)*

IS A STRONGER DOLLAR GOOD FOR THE U.S. ECONOMY?

From 2002 to 2008, the U.S. dollar lost more than a quarter of its value in foreign currency markets. On January 1, 2002, one dollar was worth 1.11 euros. On April 24, 2008 it hit its lowest point with a dollar being worth 0.64 euros. During this period, the trade deficit between the United States and the European Union grew from a yearly total of approximately −85.7 billion dollars in 2002 to 95.8

billion dollars in 2008. Was this a good thing or a bad thing for the U.S. economy?

We live in a global world. U.S. consumers buy trillions of dollars worth of imported goods and services each year, not just from the European Union, but from all over the world. U.S. businesses sell trillions of dollars' worth of exports. U.S. citizens, businesses, and governments invest trillions of dollars abroad every year. Foreign investors, businesses, and governments invest trillions of dollars in the United States each year. Indeed, foreigners are a major buyer of U.S. federal debt.

Many people feel that a weaker dollar is bad for America, that it's an indication of a weak economy. But is it? This module will help answer that question.

The world has over 150 different currencies, from the Afghanistan afghani and the Albanian lek all the way through the alphabet to the Zambian kwacha and the Zimbabwean dollar. For international economic transactions, households or firms will wish to exchange one currency for another. Perhaps the need for exchanging currencies will come from a German firm that exports products to Russia, but then wishes to exchange the Russian rubles it has earned for euros, so that the firm can pay its workers and suppliers in Germany. Perhaps it will be a South African firm that wishes to purchase a mining operation in Angola, but to make the purchase it must convert South African rand to Angolan kwanza. Perhaps it will be an American tourist visiting China, who wishes to convert U.S. dollars to Chinese yuan to pay the hotel bill.

Exchange rates can sometimes change very swiftly. For example, in the United Kingdom the pound was worth $2 in U.S. currency in spring 2008, but was worth only $1.40 in U.S. currency six months later. For firms engaged in international buying, selling, lending, and borrowing, these swings in exchange rates can have an enormous effect on profits.

This module discusses the international dimension of money, which involves conversions from one currency to another at an exchange rate. An *exchange rate* is nothing more than a price—that is, the price of one currency in terms of another currency—and so they can be analyzed with the tools of *supply* and *demand*. The first section of this module begins with an overview of foreign exchange markets: their size, their main participants, and the vocabulary for discussing movements of exchange rates. The following module uses demand and supply graphs to analyze some of the main factors that cause shifts in exchange rates. A final module then brings the central bank and monetary policy back into the picture. Each country must decide whether to allow its exchange rate to be determined in the market, or have the central bank intervene in the exchange rate market. All the choices for exchange rate policy involve distinctive tradeoffs and risks.

282. Reading: The Foreign Exchange Market

The Foreign Exchange Market

Most countries have different currencies, but not all. Sometimes small economies use the currency of an economically larger neighbor. For example, Ecuador, El Salvador, and Panama have decided to *dollarize* — that is, to use the U.S. dollar as their currency. Sometimes nations share a common currency. A large-scale example of a common currency is the decision by 17 European nations—including some very large economies such as France, Germany, and Italy—to replace their former currencies with the euro. With these exceptions duly noted, most of the international economy takes place in a situation of multiple national currencies in which both people and firms need to convert from one currency to another when selling, buying, hiring, borrowing, traveling, or investing across national borders. The market in which people or firms use one currency to purchase another currency is called the *foreign exchange market*.

You have encountered the basic concept of exchange rates in earlier modules. In The International Trade and Capital Flows, for example, we discussed how exchange rates are used to compare GDP statistics from countries where GDP is measured in different currencies. These earlier examples, however, took the actual exchange rate as given, as if it were a fact of nature. In reality, the exchange rate is a price—the price of one currency expressed in terms of units of another currency. The key framework for analyzing prices, whether in this course, any other economics course, in public policy, or business examples, is the operation of supply and demand in markets.

LINK IT UP

Visit this website for an exchange rate calculator.

The Extraordinary Size of the Foreign Exchange Markets

The quantities traded in foreign exchange markets are breathtaking. A survey done in April, 2013 by the Bank of International Settlements, an international organization for banks and the financial industry, found that $5.3 trillion *per day* was traded on foreign exchange markets, which makes the foreign exchange market the largest market in the world economy. In contrast, 2013 U.S. real GDP was $15.8 trillion *per year*.

Table 15.1 shows the currencies most commonly traded on foreign exchange markets. The foreign exchange market is dominated by the U.S. dollar, the currencies used by nations in Western Europe (the euro, the British pound, and the Australian dollar), and the Japanese yen.

Currency	% Daily Share
U.S. dollar	87.0%
Euro	33.4%
Japanese yen	23.0%
British pound	11.8%
Australian dollar	8.6%
Swiss franc	5.2%
Canadian dollar	4.6%
Mexican peso	2.5%
Chinese yuan	2.2%

Table 15.1. Currencies Traded Most on Foreign Exchange Markets as of April, 2013 (Source: http://www.bis.org/publ/rpfx13fx.pdf)

Demanders and Suppliers of Currency in Foreign Exchange Markets

In foreign exchange markets, demand and supply become closely interrelated, because a person or firm who demands one currency must at the same time supply another currency—and vice versa. To get a sense of this, it is useful to consider four groups of people or firms who participate in the market: (1) *firms* that are involved in international trade of goods and services; (2) tourists visiting other countries; (3) international investors buying ownership (or part-ownership) of a foreign firm; (4) international investors making financial investments that do not involve ownership. Let's consider these categories in turn.

Firms that buy and sell on international markets find that their costs for workers, suppliers, and investors are measured in the currency of the nation where their production occurs, but their revenues from sales are measured in the currency of the different nation where their sales happened. So, a Chinese firm exporting abroad will earn some other currency—say, U.S. dollars—but will need Chinese yuan to pay the workers, suppliers, and investors who are based in China. In the foreign exchange markets, this firm will be a supplier of U.S. dollars and a demander of Chinese yuan.

International tourists will supply their home currency to receive the currency of the country they are visiting. For example, an American tourist who is visiting China will supply U.S. dollars into the foreign exchange market and demand Chinese yuan.

Financial investments that cross international boundaries, and require exchanging currency, are often divided into two categories. *Foreign direct investment* (*FDI*) refers to purchasing a firm (at least ten percent) in another country or starting up a new enterprise in a foreign country. For example, in 2008 the Belgian beer-brewing company InBev bought the U.S. beer-maker Anheuser-Busch for $52 billion. To make this purchase of a U.S. firm, InBev would have to supply euros (the currency of Belgium) to the foreign exchange market and demand U.S. dollars.

The other kind of international financial investment, *portfolio investment*, involves a purely financial investment that does not entail any management responsibility. An example would be a U.S. financial investor who purchased bonds issued by the government of the United Kingdom, or deposited money in a British bank. To make such investments, the American investor would supply U.S. dollars in the foreign exchange market and demand British pounds.

Portfolio investment is often linked to expectations about how exchange rates will shift. Look at a U.S. financial investor who is considering purchasing bonds issued in the United Kingdom. For simplicity, ignore any interest paid by the bond (which will be small in the short run anyway) and focus on exchange rates. Say that a British pound is currently worth $1.50 in U.S. currency. However, the investor believes that in a month, the British pound will be worth $1.60 in U.S. currency. Thus, as Figure 15.2 (a) shows, this investor would change $24,000 for 16,000 British pounds. In a month, if the pound is indeed worth $1.60, then the portfolio investor can trade back to U.S. dollars at the new exchange rate, and have $25,600—a nice profit. A portfolio investor who believes that the foreign exchange rate for the pound will work in the opposite direction can also invest accordingly. Say that an investor expects that the pound, now worth $1.50 in U.S. currency, will decline to $1.40. Then, as shown in Figure 15.2 (b), that investor could start off with £20,000 in British currency (borrowing the money if necessary), convert it to $30,000 in U.S. currency, wait a month, and then convert back to approximately £21,429 in British currency—again making a nice profit. Of course, this kind of investing comes without guarantees, and an investor will suffer losses if the exchange rates do not move as predicted.

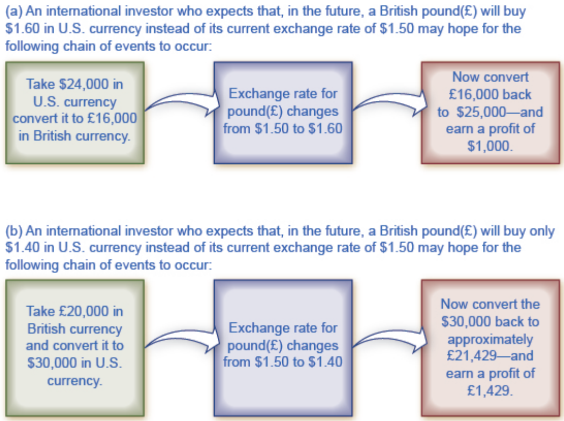


Figure 15.2. *A Portfolio Investor Trying to Benefit from Exchange Rate Movements. Expectations of the future value of a currency can drive demand and supply of that currency in foreign exchange markets.*

Many portfolio investment decisions are not as simple as betting that the value of the currency will change in one direction or the other. Instead, they involve firms trying to protect themselves from movements in exchange rates. Imagine you are running a U.S. firm that is exporting to France. You have signed a contract to deliver certain products and will receive 1 million euros a year from now.

But you do not know how much this contract will be worth in U.S. dollars, because the dollar/euro exchange rate can fluctuate in the next year. Let's say you want to know for sure what the contract will be worth, and not take a risk that the euro will be worth less in U.S. dollars than it currently is. You can *hedge*, which means using a financial transaction to protect yourself against currency risk. Specifically, you can sign a financial contract and pay a fee that guarantees you a certain exchange rate one year from now—regardless of what the market exchange rate is at that time. Now, it is possible that the euro will be worth more in dollars a year from now, so your hedging contract will be unnecessary, and you will have paid a fee for nothing. But if the value of the euro in dollars declines, then you are protected by the hedge. Financial contracts like hedging, where parties wish to be protected against exchange rate movements, also commonly lead to a series of portfolio investments by the firm that is receiving a fee to provide the hedge.

Both foreign direct investment and portfolio investment involve an investor who supplies domestic currency and demands a foreign currency. With portfolio investment less than ten percent of a company is purchased. As such, portfolio investment is often made with a short term focus. With foreign direct investment more than ten percent of a company is purchased and the investor typically assumes some managerial responsibility; thus foreign direct investment tends to have a more long-run focus. As a practical matter, portfolio investments can be withdrawn from a country much more quickly than foreign direct investments. A U.S. portfolio investor who wants to buy or sell bonds issued by the government of the United Kingdom can do so with a phone call or a few clicks of a computer key. However, a U.S. firm that wants to buy or sell a company, such as one that manufactures automobile parts in the United Kingdom, will find that planning and carrying out the transaction takes a few weeks, even months. Table 15.2 summarizes the main categories of demanders and suppliers of currency.

Table 15.2 The Demand and Supply Line-ups in Foreign Exchange Markets

Demand for the U.S. Dollar Comes from…	Supply of the U.S. Dollar Comes from…
A U.S. exporting firm that earned foreign currency and is trying to pay U.S.-based expenses	A foreign firm that has sold imported goods in the United States, earned U.S. dollars, and is trying to pay expenses incurred in its home country
Foreign tourists visiting the United States	U.S. tourists leaving to visit other countries
Foreign investors who wish to make direct investments in the U.S. economy	U.S. investors who want to make foreign direct investments in other countries
Foreign investors who wish to make portfolio investments in the U.S. economy	U.S. investors who want to make portfolio investments in other countries

Participants in the Exchange Rate Market

The foreign exchange market does not involve the ultimate suppliers and demanders of foreign exchange literally seeking each other out. If Martina decides to leave her home in Venezuela and take a trip in the United States, she does not need to find a U.S. citizen who is planning to take a vacation in Venezuela and arrange a person-to-person currency trade. Instead, the foreign exchange market works through financial institutions, and it operates on several levels.

Most people and firms who are exchanging a substantial quantity of currency go to a bank, and most banks provide foreign exchange as a service to customers. These banks (and a few other firms), known as *dealers*, then trade the foreign exchange. This is called the *interbank market*.

In the world economy, roughly 2,000 firms are foreign exchange dealers. The U.S. economy has less than 100 foreign exchange dealers, but the largest 12 or so dealers carry out more than half the total transactions. The foreign exchange market has no central location, but the major dealers keep a close watch on each other at all times.

The foreign exchange market is huge not because of the demands of tourists, firms, or even foreign direct investment, but instead because of portfolio investment and the actions of interlocking foreign exchange dealers. International tourism is a very large industry, involving about $1 trillion per year. Global exports are about 23% of global GDP; which is about $18 trillion per year. Foreign direct investment totaled about $1.4 trillion in 2012. These quantities are dwarfed, however, by the $5.3 trillion *per day* being traded in foreign exchange markets. Most transactions in the foreign exchange market are for portfolio investment—relatively short-term movements of financial capital between currencies—and because of the actions of the large foreign exchange dealers as they constantly buy and sell with each other.

283. Reading: Strengthening and Weakening Currency

Strengthening and Weakening Currency

When the prices of most goods and services change, the price is said to "rise" or "fall." For exchange rates, the terminology is different. When the exchange rate for a currency rises, so that the currency exchanges for more of other currencies, it is referred to as *appreciating* or "strengthening." When the exchange rate for a currency falls, so that a currency trades for less of other currencies, it is referred to as *depreciating* or "weakening."

To illustrate the use of these terms, consider the exchange rate between the U.S. dollar and the Canadian dollar since 1980, shown in Figure 15.3 (a). The vertical axis in Figure 15.3 (a) shows the price of $1 in U.S. currency, measured in terms of Canadian currency. Clearly, exchange rates can move up and down substantially. A U.S. dollar traded for $1.17 Canadian in 1980. The U.S. dollar appreciated or strengthened to $1.39 Canadian in 1986, depreciated or weakened to $1.15 Canadian in 1991, and then appreciated or strengthened to $1.60 Canadian by early in 2002, fell to roughly $1.20 Canadian in 2009, and then had a sharp spike up and decline in 2009 and 2010. The units in which exchange rates are measured can be confusing, because the exchange rate of the U.S. dollar is being measured using a different currency—the Canadian dollar. But exchange rates always measure the price of one unit of currency by using a different currency.

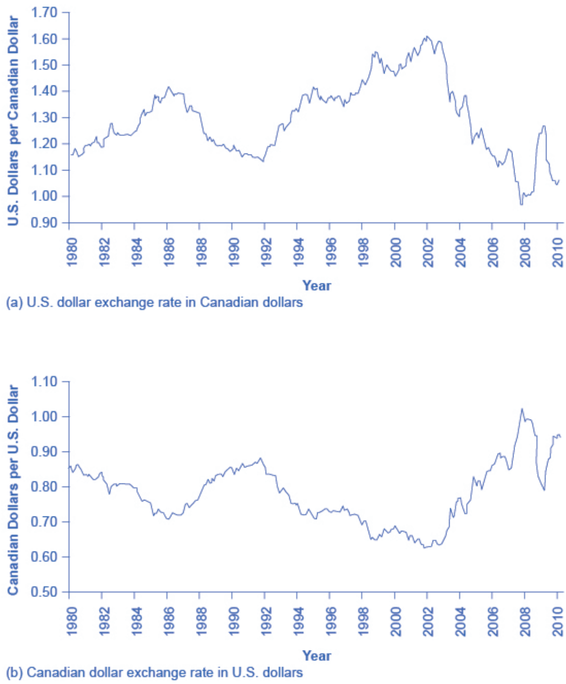


Figure 15.3. *Strengthen or Appreciate vs. Weaken or Depreciate. Exchange rates move up and down substantially, even between close neighbors like the United States and Canada. The values in (a) are a mirror image of (b); that is, any appreciation of one currency must mean depreciation of the other currency, and vice versa. (Source: http://research.stlouisfed.org/fred2/series/FXRATE…)*

In looking at the exchange rate between two currencies, the appreciation or strengthening of one currency must mean the depreciation or weakening of the other. Figure 15.3 (b) shows the exchange rate for the Canadian dollar, measured in terms of U.S. dollars. The exchange rate of the U.S. dollar measured in Canadian dollars, shown in Figure 15.3 (a), is a perfect mirror image with the exchange rate of the Canadian dollar measured in U.S. dollars, shown in Figure 15.3 (b). A fall in the Canada $/U.S. $ ratio means a rise in the U.S. $/Canada $ ratio, and vice versa.

With the price of a typical good or service, it is clear that higher prices benefit sellers and hurt buyers, while lower prices benefit buyers and hurt sellers. In the case of exchange rates, where the buyers and sellers are not always intuitively obvious, it is useful to trace through how different participants in the market will be affected by a stronger or weaker currency. Consider, for example, the impact of a stronger U.S. dollar on six different groups of economic actors, as shown in Figure 15.4: (1) U.S. exporters selling abroad; (2) foreign exporters (that is, firms selling imports in the U.S. economy); (3) U.S. tourists abroad; (4) foreign tourists visiting the United States; (5) U.S. investors (either foreign direct investment or portfolio investment) considering opportunities in other countries; (6) and foreign investors considering opportunities in the U.S. economy.

Figure 15.4. *How Do Exchange Rate Movements Affect Each Group? Exchange rate movements affect exporters, tourists, and international investors in different ways.*

For a U.S. firm selling abroad, a stronger U.S. dollar is a curse.

A strong U.S. dollar means that foreign currencies are correspondingly weak. When this exporting firm earns foreign currencies through its export sales, and then converts them back to U.S. dollars to pay workers, suppliers, and investors, the stronger dollar means that the foreign currency buys fewer U.S. dollars than if the currency had not strengthened, and that the firm's profits (as measured in dollars) fall. As a result, the firm may choose to reduce its exports, or it may raise its selling price, which will also tend to reduce its exports. In this way, a stronger currency reduces a country's exports.

Conversely, for a foreign firm selling in the U.S. economy, a stronger dollar is a blessing. Each dollar earned through export sales, when traded back into the home currency of the exporting firm, will now buy more of the home currency than expected before the dollar had strengthened. As a result, the stronger dollar means that the importing firm will earn higher profits than expected. The firm will seek to expand its sales in the U.S. economy, or it may reduce prices, which will also lead to expanded sales. In this way, a stronger U.S. dollar means that consumers will purchase more from foreign producers, expanding the country's level of imports.

For a U.S. tourist abroad, who is exchanging U.S. dollars for foreign currency as necessary, a stronger U.S. dollar is a benefit. The tourist receives more foreign currency for each U.S. dollar, and consequently the cost of the trip in U.S. dollars is lower. When a country's currency is strong, it is a good time for citizens of that country to tour abroad. Imagine a U.S. tourist who has saved up $5,000 for a trip to South Africa. In January 2008, $1 bought 7 South African rand, so the tourist had 35,000 rand to spend. In January 2009, $1 bought 10 rand, so the tourist had 50,000 rand to spend. By January 2010, $1 bought only 7.5 rand. Clearly, 2009 was the year for U.S. tourists to visit South Africa. For foreign visitors to the United States, the opposite pattern holds true. A relatively stronger U.S. dollar means that their own currencies are relatively weaker, so that as they shift from their own currency to U.S. dollars, they have fewer

U.S. dollars than previously. When a country's currency is strong, it is not an especially good time for foreign tourists to visit.

A stronger dollar injures the prospects of a U.S. financial investor who has already invested money in another country. A U.S. financial investor abroad must first convert U.S. dollars to a foreign currency, invest in a foreign country, and then later convert that foreign currency back to U.S. dollars. If in the meantime the U.S. dollar becomes stronger and the foreign currency becomes weaker, then when the investor converts back to U.S. dollars, the rate of return on that investment will be less than originally expected at the time it was made.

However, a stronger U.S. dollar boosts the returns of a foreign investor putting money into a U.S. investment. That foreign investor converts from the home currency to U.S. dollars and seeks a U.S. investment, while later planning to switch back to the home currency. If, in the meantime, the dollar grows stronger, then when the time comes to convert from U.S. dollars back to the foreign currency, the investor will receive more foreign currency than expected at the time the original investment was made.

The preceding paragraphs all focus on the case where the U.S. dollar becomes stronger. The corresponding happy or unhappy economic reactions are illustrated in the first column of Figure 15.4. The following feature centers the analysis on the opposite: a weaker dollar.

EFFECTS OF A WEAKER DOLLAR

Let's work through the effects of a weaker dollar on a U.S. exporter, a foreign exporter into the United States, a U.S. tourist going abroad, a foreign tourist coming to the United States, a U.S. investor abroad, and a foreign investor in the United States.

Step 1. Note that the demand for U.S. exports is a function of the price of those exports, which depends on the dollar price of those goods and the exchange rate of the dollar in terms of foreign currency. For example, a Ford pickup truck costs $25,000 in the United States. When it is sold in the United Kingdom, the price is $25,000 / $1.50 per British pound, or £16,667. The dollar affects the price faced by foreigners who may purchase U.S. exports.

Step 2. Consider that, if the dollar weakens, the pound rises in value. If the pound rises to $2.00 per pound, then the price of a Ford pickup is now $25,000 / $2.00 = £12,500. A weaker dollar means the foreign currency buys more dollars, which means that U.S. exports appear less expensive.

Step 3. Summarize that a weaker U.S. dollar leads to an increase in U.S. exports. For a foreign exporter, the outcome is just the opposite.

Step 4. Suppose a brewery in England is interested in selling its Bass Ale to a grocery store in the United States. If the price of a six pack of Bass Ale is £6.00 and the exchange rate is $1.50 per British pound, the price for the grocery store is 6.00 × $1.50 = $9.00 per six pack. If the dollar weakens to $2.00 per pound, the price of Bass Ale is now 6.00 × $2.00 = $12.

Step 5. Summarize that, from the perspective of U.S. purchasers, a weaker dollar means that foreign currency is more expensive, which means that foreign goods are more expensive also. This leads to a decrease in U.S. imports, which is bad for the foreign exporter.

Step 6. Consider U.S. tourists going abroad. They face the same situation as a U.S. importer—they are purchasing a foreign trip. A weaker dollar means that their trip will cost more, since a given expenditure of foreign currency (e.g., hotel bill) will take more dollars. The result is that the tourist may not stay as long abroad, and some may choose not to travel at all.

Step 7. Consider that, for the foreign tourist to the United States, a weaker dollar is a boon. It means their currency goes further, so the cost of a trip to the United States will be less. Foreigners may choose to take longer trips to the United States, and more foreign tourists may decide to take U.S. trips.

Step 8. Note that a U.S. investor abroad faces the same situation as a U.S. importer—they are purchasing a foreign asset. A U.S. investor will see a weaker dollar as an increase in the "price" of investment, since the same number of dollars will buy less foreign currency and thus less foreign assets. This should decrease the amount of U.S. investment abroad.

Step 9. Note also that foreign investors in the Unites States will have the opposite experience. Since foreign currency buys more dollars, they will likely invest in more U.S. assets.

At this point, you should have a good sense of the major players in the foreign exchange market: firms involved in international trade, tourists, international financial investors, banks, and foreign exchange dealers. The next module shows how the tools of demand and supply can be used in foreign exchange markets to explain the underlying causes of stronger and weaker currencies.

WHY IS A STRONGER CURRENCY NOT NECESSARILY BETTER?

One common misunderstanding about exchange rates is that a "stronger" or "appreciating" currency must be better than a "weaker" or "depreciating" currency. After all, is it not obvious that "strong" is better than "weak"? But do not let the terminology confuse you. When a currency becomes stronger, so that it purchases more of other currencies, it benefits some in the economy and injures others. Stronger currency is not necessarily better, it is just different.

284. Reading: Demand and Supply Shifts in Foreign Exchange Markets

Demand and Supply Shifts in Foreign Exchange Markets

The foreign exchange market involves firms, households, and investors who demand and supply currencies coming together through their banks and the key foreign exchange dealers. Figure 15.5 (a) offers an example for the exchange rate between the U.S. dollar and the Mexican peso. The vertical axis shows the exchange rate for U.S. dollars, which in this case is measured in pesos. The horizontal axis shows the quantity of U.S. dollars being traded in the foreign exchange market each day. The demand curve (D) for U.S. dollars intersects with the supply curve (S) of U.S. dollars at the equilibrium point (E), which is an exchange rate of 10 pesos per dollar and a total volume of $8.5 billion.

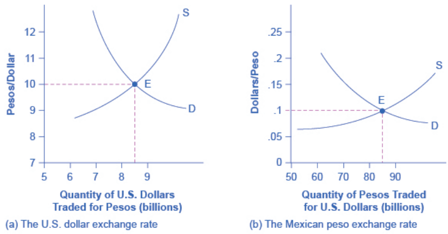


Figure 15.5. *Demand and Supply for the U.S. Dollar and Mexican Peso Exchange Rate. (a) The quantity measured on the horizontal axis is in U.S. dollars, and the exchange rate on the vertical axis is the price of U.S. dollars measured in Mexican pesos. (b) The quantity measured on the horizontal axis is in Mexican pesos, while the price on the vertical axis is the price of pesos measured in U.S. dollars. In both graphs, the equilibrium exchange rate occurs at point E, at the intersection of the demand curve (D) and the*

supply curve (S).

Figure 15.5 (b) presents the same demand and supply information from the perspective of the Mexican peso. The vertical axis shows the exchange rate for Mexican pesos, which is measured in U.S. dollars. The horizontal axis shows the quantity of Mexican pesos traded in the foreign exchange market. The *demand curve* (D) for Mexican pesos intersects with the *supply curve* (S) of Mexican pesos at the equilibrium point (E), which is an exchange rate of 10 cents in U.S. currency for each Mexican peso and a total volume of 85 billion pesos. Note that the two exchange rates are inverses: 10 pesos per dollar is the same as 10 cents per peso (or $0.10 per peso). In the actual foreign exchange market, almost all of the trading for Mexican pesos is done for U.S. dollars. What factors would cause the demand or supply to shift, thus leading to a change in the *equilibrium exchange rate*? The answer to this question is discussed in the following section.

Expectations about Future Exchange Rates

One reason to demand a currency on the foreign exchange market is the belief that the value of the currency is about to increase. One reason to supply a currency—that is, sell it on the foreign exchange market—is the expectation that the value of the currency is about to decline. For example, imagine that a leading business newspaper, like the *Wall Street Journal* or the *Financial Times*, runs an article predicting that the Mexican peso will appreciate in value. The likely effects of such an article are illustrated in Figure 15.6. Demand for the Mexican peso shifts to the right, from D0 to D1, as investors become eager to purchase pesos. Conversely, the supply of pesos shifts to the left, from S0 to S1, because investors will be less willing to give them up. The result is that the equilibrium exchange rate rises from 10 cents/peso to 12 cents/peso and the equilibrium exchange rate rises from 85 billion to 90 billion pesos as the equilibrium moves from E0 to E1.

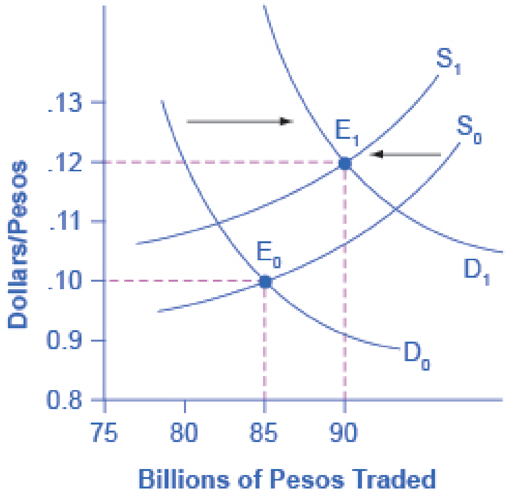


Figure 15.6. *Exchange Rate Market for Mexican Peso Reacts to Expectations about Future Exchange Rates. An announcement that the peso exchange rate is likely to strengthen in the future will lead to greater demand for the peso in the present from investors who wish to benefit from the appreciation. Similarly, it will make investors less likely to supply pesos to the foreign exchange market. Both the shift of demand to the right and the shift of supply to the left cause an immediate appreciation in the exchange rate.*

Figure 15.6 also illustrates some peculiar traits of supply and demand diagrams in the foreign exchange market. In contrast to all the other cases of supply and demand you have considered, in the *foreign exchange market*, supply and demand typically both move at the same time. Groups of participants in the foreign exchange market like firms and investors include some who are buyers and some who are sellers. An expectation of a future shift in the exchange rate affects both buyers and sellers—that is, it affects both demand and supply for a currency.

The shifts in demand and supply curves both cause the exchange rate to shift in the same direction; in this example, they both make the peso exchange rate stronger. However, the shifts in demand and supply work in opposing directions on the quantity traded. In this example, the rising demand for pesos is causing the quantity to rise while the falling supply of pesos is causing quantity to fall. In this specific example, the result is a higher quantity. But in other cases, the result could be that quantity remains unchanged or declines.

This example also helps to explain why exchange rates often move quite substantially in a short period of a few weeks or months. When investors expect a country's currency to strengthen in the future, they buy the currency and cause it to appreciate immediately. The appreciation of the currency can lead other investors to believe that future appreciation is likely—and thus lead to even further appreciation. Similarly, a fear that a currency *might* weaken quickly leads to an *actual* weakening of the currency, which often reinforces the belief that the currency is going to weaken further. Thus, beliefs about the future path of exchange rates can be self-reinforcing, at least for a time, and a large share of the trading in foreign exchange markets involves dealers trying to outguess each other on what direction exchange rates will move next.

Differences across Countries in Rates of Return

The motivation for investment, whether domestic or foreign, is to earn a return. If rates of return in a country look relatively high, then that country will tend to attract funds from abroad. Conversely, if rates of return in a country look relatively low, then funds will tend to flee to other economies. Changes in the expected rate of return will shift demand and supply for a currency. For example, imagine that interest rates rise in the United States as compared with Mexico. Thus, financial investments in the United States promise a higher return than they previously did. As a result, more investors will demand U.S. dollars so that they can buy interest-bearing assets and fewer investors will be willing to supply U.S. dollars to foreign exchange markets. Demand for the U.S. dollar will shift to the right, from D0 to D1, and supply will shift to the left, from S0 to S1, as shown in Figure 15.7. The new equilibrium (E1), will occur at an exchange rate of nine pesos/dollar and the same quantity of $8.5 billion. Thus, a higher interest rate or rate of return relative to other countries leads a nation's currency to appreciate or strengthen, and a lower interest rate relative to other countries leads a nation's currency to depreciate or weaken. Since a nation's central bank can use monetary policy to affect its interest rates, a central bank can also cause changes in exchange rates—a connection that will be discussed in more detail later in this module.

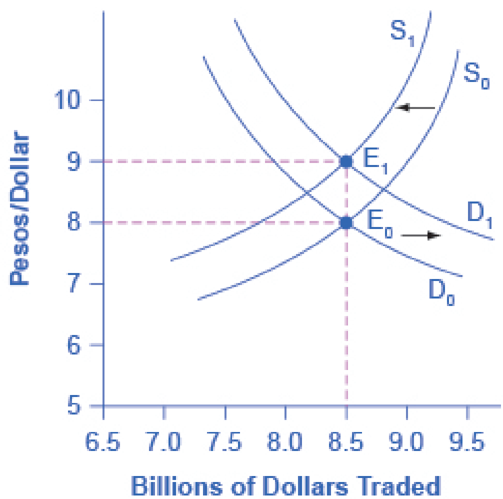


Figure 15.7. *Exchange Rate Market for U.S. Dollars Reacts to Higher Interest Rates. A higher rate of return for U.S. dollars makes holding dollars more attractive. Thus, the demand for dollars in the foreign exchange market shifts to the right, from D0 to D1, while the supply of dollars shifts to the left, from S0 to S1. The new equilibrium (E1) has a stronger exchange rate than the original equilibrium (E0), but in this example, the equilibrium quantity traded does not change.*

Relative Inflation

If a country experiences a relatively high inflation rate compared with other economies, then the buying power of its currency is eroding, which will tend to discourage anyone from wanting to acquire or to hold the currency. Figure 15.8 shows an example based on an actual episode concerning the Mexican peso. In 1986–87, Mexico experienced an inflation rate of over 200%. Not surprisingly, as inflation dramatically decreased the purchasing power of the peso in Mexico, the exchange rate value of the peso declined as well. As shown in Figure 15.8, demand for the peso on foreign exchange markets decreased from D0 to D1, while supply of the peso increased from S0 to S1. The equilibrium exchange rate fell from $2.50 per peso at the original equilibrium (E0) to $0.50 per peso at the new equilibrium (E1). In this example, the quantity of pesos traded on foreign exchange markets remained the same, even as the exchange rate shifted.

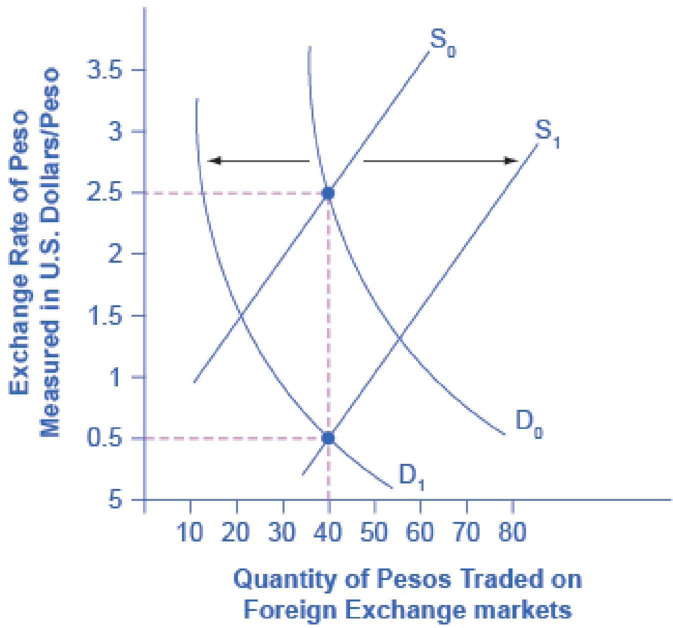


Figure 15.8. *Exchange Rate Markets React to Higher Inflation. If a currency is experiencing relatively high inflation, then its buying power is decreasing and international investors will be less eager to hold it. Thus, a rise in inflation in the Mexican peso would lead demand to shift from D0 to D1, and supply to increase from S0 to S1. Both movements in demand and supply would cause the currency to depreciate. The effect on the quantity traded is drawn here as a decrease, but in truth it could be an increase or no change, depending on the actual movements of demand and supply.*

LINK IT UP

Visit this website to learn about the Big Mac index.

Purchasing Power Parity

Over the long term, exchange rates must bear some relationship to the buying power of the currency in terms of goods that are internationally traded. If at a certain exchange rate it was much cheaper to buy internationally traded goods—such as oil, steel, computers, and cars—in one country than in another country, businesses would start buying in the cheap country, selling in other countries, and pocketing the profits.

For example, if a U.S. dollar is worth $1.60 in Canadian currency, then a car that sells for $20,000 in the United States should sell for $32,000 in Canada. If the price of cars in Canada was much lower than $32,000, then at least some U.S. car-buyers would convert their U.S. dollars to Canadian dollars and buy their cars in Canada. If the price of cars was much higher than $32,000 in this example, then at least some Canadian buyers would convert their Canadian dollars to U.S. dollars and go to the United States to purchase their cars. This is known as *arbitrage*, the process of buying and selling goods or currencies across international borders at a profit. It may occur slowly, but over time, it will force prices and exchange rates to align so that the price of internationally traded goods is similar in all countries.

The exchange rate that equalizes the prices of internationally traded goods across countries is called the *purchasing power parity* (PPP) exchange rate. A group of economists at the International Comparison Program, run by the World Bank, have calculated the PPP exchange rate for all countries, based on detailed studies of the prices and quantities of internationally tradable goods.

The purchasing power parity exchange rate has two functions. First, PPP exchange rates are often used for international comparison of GDP and other economic statistics. Imagine that you are preparing a table showing the size of *GDP* in many countries in several recent years, and for ease of comparison, you are converting all the values into U.S. dollars. When you insert the value for Japan, you need to use a yen/dollar exchange rate. But should you use the market exchange rate or the PPP exchange rate? Market exchange rates bounce around. In summer 2008, the exchange rate was 108 yen/dollar, but in late 2009 the U.S. dollar exchange rate versus the yen was 90 yen/dollar. For simplicity, say that Japan's GDP was ¥500 trillion in both 2008 and 2009. If you use the market exchange rates, then Japan's GDP will be $4.6 trillion in 2008 (that is, ¥500 trillion /(¥108/dollar)) and $5.5 trillion in 2009 (that is, ¥500 trillion /(¥90/dollar)).

Of course, it is not true that Japan's economy increased enormously in 2009—in fact, Japan had a recession like much of the rest of the world. The misleading appearance of a booming Japanese economy occurs only because we used the market exchange rate, which often has short-run rises and falls. However, PPP exchange rates stay fairly constant and change only modestly, if at all, from year to year.

The second function of PPP is that exchanges rates will often get closer and closer to it as time passes. It is true that in the short run and medium run, as exchange rates adjust to relative inflation rates, rates of return, and to expectations about how interest rates and inflation will shift, the exchange rates will often move away from the PPP exchange rate for a time. But, knowing the PPP will allow you to track and predict exchange rate relationships.

Self Check: Exchange Rates and International Finance

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the four Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/696

285. Outcome: The Balance of Trade

What you'll learn to do: explain how the balance of trade (surplus or deficit) affects the domestic economy

In this section, you will learn about how international trade is affected by fluctuations in exchange rates.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Macroeconomic Effects of Exchange Rates
- Reading: Exchange-Rate Policies
- Self Check: The Balance of Trade

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

286. Reading: Macroeconomic Effects of Exchange Rates

Exchange Rates, Aggregate Demand, and Aggregate Supply

A *central bank* will be concerned about the exchange rate for three reasons: (1) Movements in the exchange rate will affect the quantity of aggregate demand in an economy; (2) frequent substantial fluctuations in the exchange rate can disrupt international trade and cause problems in a nation's banking system; (3) the exchange rate may contribute to an unsustainable balance of trade and large inflows of international financial capital, which can set the economy up for a deep recession if international investors decide to move their money to another country. Let's discuss these scenarios in turn.

Foreign trade in goods and services typically involves incurring the costs of production in one currency while receiving revenues from sales in another currency. As a result, movements in exchange rates can have a powerful effect on incentives to export and import, and thus on *aggregate demand* in the economy as a whole.

For example, in 1999, when the euro first became a currency, its value measured in U.S. currency was $1.06/euro. By the end of 2013, the euro had risen (and the U.S. dollar had correspondingly weakened) to $1.37/euro. Consider the situation of a French firm that each year incurs €10 million in costs, and sells its products in the United States for $10 million. In 1999, when this firm converted $10 million back to euros at the exchange rate of $1.06/euro (that is, $10 million × [€1/$1.06]), it received €9.4 million, and suffered a loss. In 2013, when this same firm converted $10 million back to euros at the exchange rate of $1.37/euro (that is, $10 million ×

[€1 euro/$1.37]), it received approximately €7.3 million and an even larger loss. This example shows how a stronger euro discourages exports by the French firm, because it makes the costs of production in the domestic currency higher relative to the sales revenues earned in another country. From the point of view of the U.S. economy, the example also shows how a weaker U.S. dollar encourages *exports*.

Since an increase in exports results in more dollars flowing into the economy, and an increase in imports means more dollars are flowing out, it is easy to conclude that exports are "good" for the economy and imports are "bad," but this overlooks the role of exchange rates. If an American consumer buys a Japanese car for $20,000 instead of an American car for $30,000, it may be tempting to argue that the American economy has lost out. However, the Japanese company will have to convert those dollars to yen to pay its workers and operate its factories. Whoever buys those dollars will have to use them to purchase American goods and services, so the money comes right back into the American economy. At the same time, the consumer saves money by buying a less expensive import, and can use the extra money for other purposes.

Fluctuations in Exchange Rates

Exchange rates can fluctuate a great deal in the short run. As yet one more example, the Indian rupee moved from 39 rupees/dollar in February 2008 to 51 rupees/dollar in March 2009, a decline of more than one-fourth in the value of the rupee on foreign exchange markets. Figure 15.9 earlier showed that even two economically developed neighboring economies like the United States and Canada can see significant movements in exchange rates over a few years. For firms that depend on export sales, or firms that rely on imported inputs to production, or even purely domestic firms that compete with firms tied into international trade—which in many countries adds up to half or more of a nation's GDP—sharp movements in exchange rates can lead to dramatic changes in profits and losses. So, a central bank may desire to keep exchange rates from moving too much as part of providing a stable business climate, where firms can focus on productivity and innovation, not on reacting to exchange rate fluctuations.

One of the most economically destructive effects of exchange rate fluctuations can happen through the banking system. Most international loans are measured in a few large currencies, like U.S. dollars, European euros, and Japanese yen. In countries that do not use these currencies, banks often borrow funds in the currencies of other countries, like U.S. dollars, but then lend in their own domestic currency. The left-hand chain of events in Figure 15.9 shows how this pattern of international borrowing can work. A bank in Thailand borrows one million in U.S. dollars. Then the bank converts the dollars to its domestic currency—in the case of Thailand, the currency is the baht—at a rate of 40 baht/dollar. The bank then lends the baht to a firm in Thailand. The business repays the loan in baht, and the bank converts it back to U.S. dollars to pay off its original U.S. dollar loan.

Figure 15.9. *International Borrowing. The scenario of international borrowing that ends on the left is a success story, but the scenario that ends on the right shows what happens when the exchange rate weakens.*

This process of borrowing in a foreign currency and lending in a domestic currency can work just fine, as long as the exchange rate does not shift. In the scenario outlined, if the dollar strengthens and the baht weakens, a problem arises. The right-hand chain of events in Figure 15.9 illustrates what happens when the baht unexpectedly weakens from 40 baht/dollar to 50 baht/dollar. The Thai firm still repays the loan in full to the bank. But because of the shift in the exchange rate, the bank cannot repay its loan in U.S. dollars. (Of course, if the exchange rate had changed in the other direction, making the Thai currency stronger, the bank could have realized an unexpectedly large profit.)

In 1997–1998, countries across eastern Asia, like Thailand, Korea, Malaysia, and Indonesia, experienced a sharp depreciation of their currencies, in some cases 50% or more. These countries had been experiencing substantial inflows of *foreign investment capital*, with bank lending increasing by 20% to 30% per year through the mid-1990s. When their exchange rates depreciated, the banking systems in these countries were bankrupt. Argentina experienced a similar chain of events in 2002. When the Argentine peso depreciated, Argentina's banks found themselves unable to pay back what they had borrowed in U.S. dollars.

Banks play a vital role in any economy in facilitating transactions and in making loans to firms and consumers. When most of a country's largest banks become *bankrupt* simultaneously, a sharp decline in aggregate demand and a deep recession results. Since the main responsibilities of a central bank are to control the money supply and to ensure that the banking system is stable, a central bank must be concerned about whether large and unexpected exchange rate depreciation will drive most of the country's existing banks into bankruptcy.

Summing Up Public Policy and Exchange Rates

Every nation would prefer a stable exchange rate to facilitate *international trade* and reduce the degree of risk and uncertainty in the economy. However, a nation may sometimes want a weaker exchange rate to stimulate aggregate demand and reduce a recession, or a stronger exchange rate to fight inflation. The country must also be concerned that rapid movements from a weak to a strong exchange rate may cripple its export industries, while rapid movements from a strong to a weak exchange rate can cripple its banking sector. In short, every choice of an exchange rate—whether it should be stronger or weaker, or fixed or changing—represents potential tradeoffs.

287. Reading: Exchange-Rate Policies

Exchange-Rate Policies

Exchange rate policies come in a range of different forms listed in Figure 15.10: let the foreign exchange market determine the exchange rate; let the market set the value of the exchange rate most of the time, but have the central bank sometimes intervene to prevent fluctuations that seem too large; have the central bank guarantee a specific exchange rate; or share a currency with other countries. Let's discuss each type of exchange rate policy and its tradeoffs.

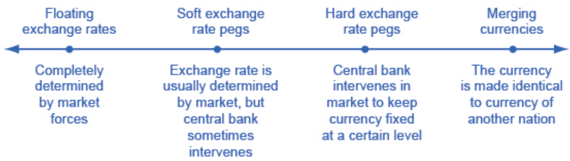


Figure 15.10. *A Spectrum of Exchange Rate Policies. A nation may adopt one of a variety of exchange rate regimes, from floating rates in which the foreign exchange market determines the rates to pegged rates where governments intervene to manage the value of the exchange rate, to a common currency where the nation adopts the currency of another country or group of countries.*

Floating Exchange Rates

A policy which allows the foreign exchange market to set exchange rates is referred to as a *floating exchange rate*. The U.S. dollar is a floating exchange rate, as are the currencies of about 40% of the countries in the world economy. The major concern with this policy is that exchange rates can move a great deal in a short time.

Consider the U.S. exchange rate expressed in terms of another fairly stable currency, the Japanese yen, as shown in Figure 15.11. On January 1, 2002, the exchange rate was 133 yen/dollar. On January 1, 2005, it was 103 yen/dollar. On June 1, 2007, it was 122 yen/dollar, and on January 1, 2009, it was 90 yen/dollar. As investor sentiment swings back and forth, driving exchange rates up and down, exporters, importers, and banks involved in international lending are all affected. At worst, large movements in exchange rates can drive companies into bankruptcy or trigger a nationwide banking collapse. But even in the moderate case of the yen/dollar exchange rate, these movements of roughly 30 percent back and forth impose stress on both economies as firms must alter their export and import plans to take the new exchange rates into account. Especially in smaller countries where international trade is a relatively large share of GDP, exchange rate movements can rattle their economies.

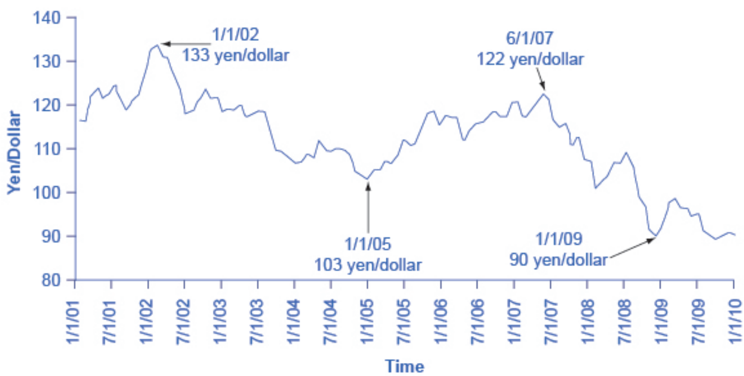


***Figure 15.11**. U.S. Dollar Exchange Rate in Japanese Yen. Even relatively stable exchange rates can vary a fair amount. The exchange rate for the U.S. dollar, measured in Japanese yen, fell about 30% from the start of 2002 to the start of 2005, rose back by mid-2007, and then dropped again by early 2009. (Source: http://research.stlouisfed.org/fred2/series/EXJPUS)*

However, movements of floating exchange rates have advantages, too. After all, prices of goods and services rise and fall throughout a market economy, as demand and supply shift. If an economy experiences strong inflows or outflows of international financial capital, or has relatively high inflation, or if it experiences strong productivity growth so that purchasing power changes relative to other economies, then it makes economic sense for the exchange rate to shift as well.

Floating exchange rate advocates often argue that if government policies were more predictable and stable, then inflation rates and interest rates would be more predictable and stable. Exchange rates would bounce around less, too. The great economist *Milton Friedman* (1912–2006), for example, wrote a defense of floating exchange rates in 1962 in his book *Capitalism and Freedom*:

> Being in favor of floating exchange rates does not mean being in favor of unstable exchange rates. When we support a free price system [for goods and services] at home, this does not imply that we favor a system in which prices fluctuate wildly up and down. What we want is a system in which prices are free to fluctuate but in which the forces determining them are sufficiently stable so that in fact prices move within moderate ranges. This is equally true in a system of floating exchange rates. The ultimate objective is a world in which exchange rates, while free to vary, are, in fact, highly stable because basic economic policies and conditions are stable.

Advocates of floating exchange rates admit that, yes, exchange rates may sometimes fluctuate. They point out, however, that if a central bank focuses on preventing either high inflation or deep recession, with low and reasonably steady interest rates, then exchange rates will have less reason to vary.

Using Soft Pegs and Hard Pegs

When a government intervenes in the foreign exchange market so that the exchange rate of its currency is different from what the market would have produced, it is said to have established a "peg" for its currency. A *soft peg* is the name for an exchange rate policy where the government usually allows the exchange rate to be set by the market, but in some cases, especially if the exchange rate seems to be moving rapidly in one direction, the central bank will intervene in the market. With a *hard peg* exchange rate policy, the central bank sets a fixed and unchanging value for the exchange rate. A central bank can implement soft peg and hard peg policies.

Suppose the market exchange rate for the Brazilian currency, the real, would be 35 cents/real with a daily quantity of 15 billion real traded in the market, as shown at the equilibrium E0 in Figure 15.12 (a) and Figure 15.12 (b). However, the government of Brazil decides that the exchange rate should be 30 cents/real, as shown in Figure 15.12 (a). Perhaps Brazil sets this lower exchange rate to benefit its export industries. Perhaps it is an attempt to stimulate aggregate demand by stimulating exports. Perhaps Brazil believes that the current market exchange rate is higher than the long-term purchasing power parity value of the real, so it is minimizing fluctuations in the real by keeping it at this lower rate. Perhaps the target exchange rate was set sometime in the past, and is now being maintained for the sake of stability. Whatever the reason, if Brazil's central bank wishes to keep the exchange rate below the market level, it must face the reality that at this weaker exchange rate of 30 cents/real, the quantity demanded of its currency at 17 billion reals is greater than the quantity supplied of 13 billion reals in the foreign exchange market.

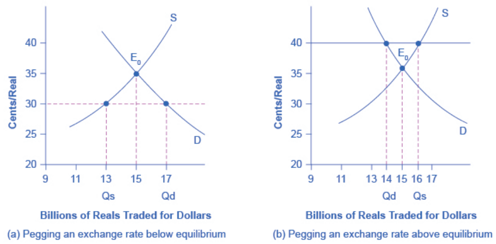


Figure 15.12. *Pegging an Exchange Rate. (a) If an exchange rate is pegged below what would otherwise be the equilibrium, then the quantity demanded of the currency will exceed the quantity supplied. (b) If an exchange rate is pegged above what would otherwise be the equilibrium, then the quantity supplied of the currency exceeds the quantity demanded.*

The Brazilian central bank could weaken its exchange rate in two ways. One approach is to use an *expansionary monetary policy* that leads to lower interest rates. In foreign exchange markets, the lower interest rates will reduce demand and increase supply of the real and lead to depreciation. This technique is not often used because lowering interest rates to weaken the currency may be in conflict with the country's monetary policy goals. Alternatively, Brazil's central bank could trade directly in the foreign exchange market. The central bank can expand the money supply by creating reals, use the reals to purchase foreign currencies, and avoid selling any of its own currency. In this way, it can fill the gap between *quantity demanded* and *quantity supplied* of its currency.

Figure 15.12 (b) shows the opposite situation. Here, the Brazilian government desires a stronger exchange rate of 40 cents/real than the market rate of 35 cents/real. Perhaps Brazil desires the stronger currency to reduce aggregate demand and to fight inflation, or perhaps Brazil believes that that current market exchange rate is temporarily lower than the long-term rate. Whatever the reason, at the higher desired exchange rate, the quantity supplied of 16 billion reals exceeds the quantity demanded of 14 billion reals.

Brazil's central bank can use a *contractionary monetary policy* to raise interest rates, which will increase demand and reduce supply of the currency on foreign exchange markets, and lead to an appreciation. Alternatively, Brazil's central bank can trade directly in the foreign exchange market. In this case, with an excess supply of its own currency in foreign exchange markets, the central bank must use reserves of foreign currency, like U.S. dollars, to demand its own currency and thus cause an appreciation of its exchange rate.

Both a soft peg and a hard peg policy require that the central bank intervene in the foreign exchange market. However, a hard peg policy attempts to preserve a fixed exchange rate at all times. A soft peg policy typically allows the exchange rate to move up and down by relatively small amounts in the short run of several months or a year, and to move by larger amounts over time, but seeks to avoid extreme short-term fluctuations.

Tradeoffs of Soft Pegs and Hard Pegs

When a country decides to alter the market exchange rate, it faces a number of tradeoffs. If it uses *monetary policy* to alter the exchange rate, it then cannot at the same time use monetary policy to address issues of inflation or recession. If it uses direct purchases and sales of foreign currencies in exchange rates, then it must face the issue of how it will handle its reserves of foreign currency. Finally, a pegged exchange rate can even create additional movements of the exchange rate; for example, even the possibility of government intervention in exchange rate markets will lead to rumors about whether and when the government will intervene, and dealers in the foreign exchange market will react to those rumors. Let's consider these issues in turn.

One concern with pegged exchange rate policies is that they imply a country's monetary policy is no longer focused on controlling inflation or shortening recessions, but now must also take the exchange rate into account. For example, when a country pegs its exchange rate, it will sometimes face economic situations where it would like to have an expansionary monetary policy to fight recession—but it cannot do so because that policy would depreciate its exchange rate and break its hard peg. With a soft peg exchange rate policy, the *central bank* can sometimes ignore the exchange rate and focus on domestic inflation or recession—but in other cases the central bank may ignore inflation or recession and instead focus on its soft peg exchange rate. With a hard peg policy, domestic monetary policy is effectively no longer determined by domestic inflation or unemployment, but only by what monetary policy is needed to keep the exchange rate at the hard peg.

Another issue arises when a central bank intervenes directly in the exchange rate market. If a central bank ends up in a situation where it is perpetually creating and selling its own currency on foreign exchange markets, it will be buying the currency of other countries, like U.S. dollars or euros, to hold as *reserves*. Holding large reserves of other currencies has an *opportunity cost*, and central banks will not wish to boost such reserves without limit.

In addition, a central bank that causes a large increase in the supply of money is also risking an inflationary surge in aggregate demand. Conversely, when a central bank wishes to buy its own currency, it can do so by using its reserves of international currency like the U.S. dollar or the euro. But if the central bank runs out of such reserves, it can no longer use this method to strengthen its currency. Thus, buying foreign currencies in exchange rate markets can be expensive and inflationary, while selling foreign currencies can work only until a central bank runs out of reserves.

Yet another issue is that when a government pegs its exchange rate, it may unintentionally create another reason for additional fluctuation. With a soft peg policy, foreign exchange dealers and international investors react to every rumor about how or when the central bank is likely to intervene to influence the exchange rate, and as they react to rumors the exchange rate will shift up and down. Thus, even though the goal of a soft peg policy is to reduce short-term fluctuations of the exchange rate, the existence of the policy—when anticipated in the foreign exchange market—may sometimes increase short-term fluctuations as international investors try to anticipate how and when the central bank will act. The following section discusses the effects of *international capital flows*—capital that flows across national boundaries as either portfolio investment or direct investment.

HOW DO TOBIN TAXES CONTROL THE FLOW OF CAPITAL?

Some countries like Chile and Malaysia have sought to reduce movements in exchange rates by limiting inflows and outflows of international financial capital. This policy can be enacted either through targeted taxes or by regulations.

Taxes on international capital flows are sometimes known as *Tobin taxes*, named after *James Tobin*, the 1981 Nobel laureate in economics who proposed such a tax in a 1972 lecture. For example, a government might tax all foreign exchange transactions, or attempt to tax short-term *portfolio investment* while exempting long-term foreign *direct investment*. Countries can also use regulation to forbid certain kinds of foreign investment in the first place or to make it difficult for international financial investors to withdraw their funds from a country.

The goal of such policies is to reduce international capital flows, especially short-term portfolio flows, in the hope that doing so will reduce the chance of large movements in exchange rates that can bring macroeconomic disaster.

But proposals to limit *international financial flows* have severe practical difficulties. Taxes are imposed by national governments, not international ones. If one government imposes a Tobin tax on exchange rate transactions carried out within its territory, the exchange rate market might easily be operated by a firm based someplace like the Grand Caymans, an island nation in the Caribbean well-known for allowing some financial wheeling and dealing. In an interconnected global economy, if goods and services are allowed to flow across national borders, then payments need to flow across borders, too. It is very difficult—in fact close to impossible—for a nation to allow only the flows of payments that relate to goods and services, while clamping down or taxing other flows of financial capital. If a nation participates in international trade, it must also participate in international capital movements.

Finally, countries all over the world, especially low-income countries, are crying out for foreign investment to help develop their economies. Policies that discourage international financial investment may prevent some possible harm, but they rule out potentially substantial economic benefits as well.

A hard peg exchange rate policy will not allow short-term fluctuations in the exchange rate. If the government first announces a hard peg and then later changes its mind—perhaps the government becomes unwilling to keep interest rates high or to hold high levels of foreign exchange reserves—then the result of abandoning a hard peg could be a dramatic shift in the exchange rate.

In the mid-2000s, about one-third of the countries in the world used a soft peg approach and about one-quarter used a hard peg approach. The general trend in the 1990s was to shift away from a soft peg approach in favor of either floating rates or a hard peg. The concern is that a successful soft peg policy may, for a time, lead to very little variation in exchange rates, so that firms and banks in the economy begin to act as if a hard peg exists. When the exchange rate does move, the effects are especially painful because firms and banks have not planned and hedged against a possible change. Thus, the argument went, it is better either to be clear that the exchange rate is always flexible, or that it is fixed, but choosing an in-between soft peg option may end up being worst of all.

A Merged Currency

A final approach to exchange rate policy is for a nation to choose a common currency shared with one or more nations is also called a *merged currency*. A merged currency approach eliminates foreign exchange risk altogether. Just as no one worries about exchange rate movements when buying and selling between New York and California, Europeans know that the value of the euro will be the same in Germany and France and other European nations that have adopted the euro.

However, a merged currency also poses problems. Like a hard peg, a merged currency means that a nation has given up altogether on domestic monetary policy, and instead has put its interest rate policies in other hands. When Ecuador uses the U.S. dollar as its currency, it has no voice in whether the Federal Reserve raises or lowers interest rates. The European Central Bank that determines monetary policy for the euro has representatives from all the euro nations. However, from the standpoint of, say, Portugal, there will be times when the decisions of the European Central Bank about monetary policy do not match the decisions that would have been made by a Portuguese central bank.

The lines between these four different exchange rate policies can blend into each other. For example, a soft peg exchange rate policy in which the government almost never acts to intervene in the exchange rate market will look a great deal like a floating exchange rate. Conversely, a soft peg policy in which the government intervenes often to keep the exchange rate near a specific level will look a lot like a hard peg. A decision to merge currencies with another country is, in effect, a decision to have a permanently fixed exchange rate with those countries, which is like a very hard exchange rate peg. The range of exchange rates policy choices, with their advantages and disadvantages, are summarized in Table 15.3.

Table 15.3. Tradeoffs of Exchange Rate Policies

Situation	Floating Exchange Rates	Soft Peg	Hard Peg	Merged Currency
Large short-run fluctuations in exchange rates?	Often a lot in the short term	Maybe less in the short run, but still large changes over time	None, unless a change in the fixed rate	None
Large long-term fluctuations in exchange rates?	Can often happen	Can often happen	Cannot happen unless hard peg changes, in which case substantial volatility can occur	Cannot happen
Power of central bank to conduct countercyclical monetary policy?	Flexible exchange rates make monetary policy stronger	Some power, although conflicts may arise between exchange rate policy and countercyclical policy	Very little; central bank must keep exchange rate fixed	None; nation does not have its own currency
Costs of holding foreign exchange reserves?	Do not need to hold reserves	Hold moderate reserves that rise and fall over time	Hold large reserves	No need to hold reserves
Risk of being stuck with an exchange rate that causes a large trade imbalance and very high inflows or outflows of financial capital?	Adjusts often	Adjusts over the medium term, if not the short term	May become stuck over time either far above or below the market level	Cannot adjust

Global macroeconomics would be easier if the whole world had one currency and one central bank. The exchange rates between different currencies complicate the picture. If exchange rates are set solely by financial markets, they fluctuate substantially as short-term portfolio investors try to anticipate tomorrow's news. If the government attempts to intervene in exchange rate markets through soft pegs or hard pegs, it gives up at least some of the power to use monetary policy to focus on domestic inflations and recessions, and it risks causing even greater fluctuations in foreign exchange markets.

There is no consensus among economists about which exchange rate policies are best: floating, soft peg, hard peg, or merged currencies. The choice depends both on how well a nation's central bank can implement a specific exchange rate policy and on how well a nation's firms and banks can adapt to different exchange rate policies. A national economy that does a fairly good job at achieving the four main economic goals of growth, low inflation, low unemployment, and a sustainable *balance of trade* will probably do just fine most of the time with any exchange rate policy; conversely, no exchange rate policy is likely to save an economy that consistently fails at achieving these goals. On the other hand, a merged currency applied across wide geographic and cultural areas carries with it its own set of problems, such as the ability for countries to conduct their own independent monetary policies.

IS A STRONGER DOLLAR GOOD FOR THE U.S. ECONOMY?

The foreign exchange value of the dollar is a price and whether a higher price is good or bad depends on where you are standing: sellers benefit from higher prices and buyers are harmed. A stronger dollar is good for U.S. imports (and people working for U.S. importers) and U.S. investment abroad. It is also good for U.S. tourists going to other countries, since their dollar goes further. But a stronger dollar is bad for U.S. exports (and people working in U.S. export industries); it is bad for foreign investment in the United States (leading, for example, to higher U.S. interest rates); and it is bad for foreign tourists (as well as U.S hotels, restaurants, and others in the tourist industry). In short, whether the U.S. dollar is good or bad is a more complex question than you may have thought. The economic answer is “it depends.”

Self Check: The Balance of Trade

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You’ll have more success on the Self Check if you’ve completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/697

288. Outcome: Globalization

What you'll learn to do: connect globalization, international trade and international finance

Trade and Finance are often confused as being synonymous with globalization. Indeed, trade and international finance have contributed to globalization but they are not the same. Globalization is a process that widens, deepens and speeds-up interconnectedness between people, institutions, markets and nations. Trade and finance are two arteries through which the process of globalization flows. There are many ways to conceptualize globalization.

As globalization increases over time, individuals, firms, institutions, and politicians work within and across countries to define exactly how "open" they want to be and whether they prefer to protect their own products through tariffs or trade restrictions. One thing is for sure, protection and openness to international trade OR finance both have income distribution effects. The interesting question is who reaps the benefits and who carries the burden.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Introduction to Globalization
- Reading: Trade Winds
- Simulation: International Trade
- Self Check: Globalization

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

289. Reading: Introduction to Globalization

Conceptualizing Globalization

Globalization is the process by which the world, previously isolated through physical and technological distance, becomes increasingly interconnected. It is manifested by the increase in interaction between peoples around the world that involves the sharing of ideas, cultures, goods, services and investment.

The last sixty years have witnessed a huge increase in globalization, but the phenomenon has been going on for much longer. Thomas Friedman describes the current trend as the third great wave of globalization in human history.

Globalization has brought fear of loss of jobs and loss of income, which are often described as the "race to the bottom," as industrialized countries are thought to have to reduce wages to be competitive with those in the developing world. Globalization has also spawned fears about loss of culture. Many countries worry about their cultures being overwhelmed by that of the United States. France is a good example. Others fear replacement of their cultures by that of Western nations (e.g., some Islamic states). Countries also fear the loss of national sovereignty as they become part of supranational entitles, like the European Union or the International Monetary Fund. And yet, history shows that globalization has corresponded to higher national incomes and increased opportunities. How can these conflicting views be reconciled?

290. Reading: Trade Winds

Trade Winds

Rapid increases in the flow of goods and services between vastly different nations and cultures have changed what people eat, how they dress, and even how they communicate with one another. For you, increased trade has meant greater choice of what to buy and often lower prices.

Look through your room. Chances are it is full of items from all around the world. The relatively free trade that exists today provides you with expanded choices. No one forced you to buy that shirt from India or that e-book reader whose components are manufactured in various countries in Asia. Presumably you bought them because you preferred them to other shirts and e-book readers you might have bought, perhaps because they had certain characteristics—style, color, perceived quality, or price—that you favored.

Your gains are being experienced worldwide because the winds of international trade have blown generally freer in the past few decades. Nations all over the world have dramatically lowered the barriers they impose on the products of other countries.

In Western Europe, the members of the European Union (EU) have eliminated virtually every restriction on the free flow of goods and services among them. A truckload of electronic equipment from Italy now passes through France on its way to Spain with no more restrictions than would be encountered by a truck delivering goods from Michigan to Illinois. The purchase of the equipment can even be arranged using a new currency, the euro, which has been adopted by most EU nations.

Canada, Mexico, and the United States, while not adopting a common currency, have created a similar free trade area, the North

American Free Trade Area (NAFTA). NAFTA has resulted in a dramatic increase in trade between Canada, the United States, and Mexico.

President Bush proposed and Congress passed in 2005 the creation of a Central American Free Trade Association (CAFTA) that would create a free trade area south of Mexico and linked to the United States. It abolished most tariff restrictions between the United States and six countries of Central America—Costa Rica, the Dominican Republic, El Salvador, Guatemala, Honduras, and Nicaragua. President Bush also proposed free trade agreements with Peru, Colombia, Panama, and South Korea. The agreement with Peru passed at the end of 2007. Free trade agreements with the other three countries finally passed under the administration of President Obama in late 2011.

While many other bilateral and regional free trade agreements have gone into effect in countries around the world, a major worldwide proposal has faltered. In 1995, the World Trade Organization (WTO) was established to "help trade flow smoothly, freely, fairly and predictably" among member nations. Since 2008, it has had 153 member countries. Since World War II, the General Agreement on Tariffs and Trade (GATT)—WTO's predecessor—and WTO have generated a series of agreements that slashed trade restraints among members. These agreements have helped propel international trade, but the negotiations leading to these agreements have always been protracted and tumultuous and issues of nationalism and patriotism are often not far from the surface. The current and ninth round of trade talks are referred to as the Doha Round, because they were officially launched in Doha, Qatar, in 2001. Ten years later, talks were still mired in controversy over the removal of agricultural export subsidies and lowering of trade barriers of various kinds.

Still trade is more extensive and much freer than it was fifty or one hundred years ago. The 2008 financial crisis severely tested the attitudes of many countries toward working toward and implementing agreements that lead to freer trade. Some countries did take protectionist measures, but by and large they held to their trading system commitments and did not adopt import restrictions. A 2010 study by Hiau Looi Kee, Cristina Neagu, and Alessandro Nicita for the World Bank found that while many countries did adjust their tariffs upward or impose other restrictions on selective products, most changes were on products that did not have a significant effect on trade flows. They estimated that these protectionist policies explain less than 2% of the decline in world trade that occurred during the crisis period.

Why have so many countries moved to make trade freer? What are the effects of free trade? Why do efforts to eliminate trade restrictions meet with resistance? Why do many nations continue to impose barriers against some foreign goods and services? How do such barriers affect the economy? How do such barriers affect *you*?

Self Check: Globalization

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/698

291. Simulation: International Trade

Try It

Play the simulation below multiple times to see how different choices lead to different outcomes. All simulations allow unlimited attempts so that you can gain experience applying the concepts.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=323

292. Self Check: Globalization

Check Your Understanding

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the two Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/698

293. Putting It Together: Globalization, Trade and Finance

Summary

The goal of this module was to teach you to analyze the benefits and costs of international trade, and to determine the extent to which barriers to international trade are warranted.

You learned how to:

- Define and calculate comparative and absolute advantage
- Define and calculate gains from trade
- Understand the way imports and exports impact different actors in the economy (businesses, consumers, and workers)
- Explain how globalization has increased over time, especially over the last several decades
- Understand the way government regulations (e.g. tariffs, quotas and non-tariff barriers) affect business, consumers and workers in the economy.
- Differentiate between alternative international trade regimes and how they impact global trade
- Explain how changes in currency exchange rates impact trade balances
- Explain how the balance of trade (surplus or deficit) affects the domestic economy, and how the domestic economy affects the balance of trade

Examples

You learned that trade based on comparative advantage will maximize an individual's or a nation's income, but that there will be winners and losers to trade. For example, employees and owners of a firm that loses business to foreign imports are worse off, even though their loss is less than the gain to consumers. The challenge for policymakers is how to compensate the losers while capturing the gains from trade. Similarly, protectionism benefits some workers and businesses at the expense of other workers and businesses and at the expense of consumers. In this case, the losses to the latter groups are larger than the gains to the former groups. In that sense, protectionism makes a country worse off. Finally, a trade deficit means that a nation is consuming beyond its income, in other words, it is borrowing from the rest of the world. This implies short term benefits and long term costs as those debts must repaid. Whether or not a trade deficit makes sense depends on what is done with the borrowed resources. Borrowing to invest in the future, say for example by building railroads as the U.S. did in the late 1800s, raises the nation's future GDP and enhances their ability to pay back their loans. Borrowing to raise the current standard of living or to allow citizens to retire early, as Greece did in the early 2000s, has the opposite effect.

294. Glossary: International Trade

absolute advantage
: when one country can use fewer resources to produce a good compared to another country; when a country is more productive compared to another country

common market
: economic agreement between countries to allow free trade in goods, services, labor, and financial capital between members while having a common external trade policy

comparative advantage
: when a country can produce a good at a lower cost in terms of other goods; or, when a country has a lower opportunity cost of production

economic union
: economic agreement between countries to allow free trade between members, a common external trade policy, and coordinated monetary and fiscal policies

free trade agreement
: economic agreement between countries to allow free trade between members

gain from trade
: a country that can consume more than it can produce as a result of specialization and trade

General Agreement on Tariffs and Trade (GATT)
: forum in which nations could come together to negotiate reductions in tariffs and other barriers to trade; the precursor to the World Trade Organization

import quotas
: numerical limits on the quantity of products that can be imported

intra-industry trade
: international trade of goods within the same industry

national interest argument
: the argument that there are compelling national interests against depending on key imports from other nations

nontariff barriers
: ways a nation can draw up rules, regulations, inspections, and paperwork to make it more costly or difficult to import products

protectionism
: government policies to reduce or block imports

splitting up the value chain
: many of the different stages of producing a good happen in different geographic locations

tariffs
: taxes that governments place on imported goods

value chain
: how a good is produced in stages

World Trade Organization (WTO)
: organization that seeks to negotiate reductions in barriers to trade and to adjudicate complaints about violations of international trade policy; successor to the General Agreement on Tariffs and Trade (GATT)

295. Discussion: Absolute and Comparative Advantage

Suppose that the United States and Canada can each produce two products: lumber and beef. Create a table like the one below, showing labor requirements per unit of output for each country. (Hint: Choose numbers for each country that are easily divisible by one another.)

Labor Requirements per Unit of Output

	United States	Canada
Lumber		
Beef		

What does absolute advantage mean? How do you calculate absolute advantage? In what output(s) does the U.S. have an absolute advantage? Explain using the data from your table. In what output(s) does Canada have an absolute advantage? Explain using the data from your table.

What does comparative advantage mean? How do you calculate comparative advantage? In what output(s) does the U.S. have a comparative advantage? Explain using the data from your table. In what output(s) does Canada have a comparative advantage? Explain using the data from your table.

What product should each country export? Why?

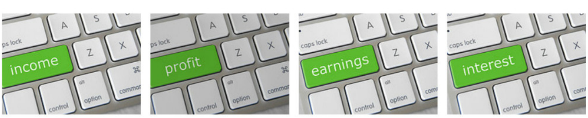
PART XVI

MODULE: INCOME DISTRIBUTION

296. Why It Matters: Income Distribution

Why assess how resource markets/factors of production affect society's distribution of income?

We have spent a great deal of time exploring product markets, that is, markets for goods and services. It's now time to look at the other side of the production process, namely, the input side. This is usually fairly interesting to students as it explains the sources of peoples' incomes. The U.S. is largely a market economy, where income is derived from ownership of resources. Owners of resources (or inputs) like labor, land and capital are paid what the market says the resources are worth. Of course, there is more than one market for labor and other inputs. For example, the market for new college graduates in economics is different from the market for Ph.D. economists, and the value markets place on new college graduates versus economists reflects that difference. But the principles of supply and demand still apply for all inputs, as we will see in this module.



Income, Earnings, Profit, Interest *by* GotCredit, CC-BY.

As you work through this module, consider the following questions:

- What factors determine how much a person gets paid for their labor?
- Why do teachers and nurses get paid less than professional athletes?
- Why do garbage men (or urban sanitation engineers) get paid significantly above minimum wage? Is it because of the skills they possess?
- Why do unionized workers earn more than non-union workers?
- How much of the income differences between people is due to chance and how much is due to choices the individuals make?
- Can poverty be eliminated or will some part of the population always be poor?
- How well does the marginal productivity theory of income distribution explain the actual income distribution in the United States or other nations?

Some of these questions will be explicitly answered in the module; others you will have to think about.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=329

LEARNING OUTCOMES

- Describe the incomes earned by the factors of production (land, labor, capital, entrepreneurship) wages, interest, rents, and profit
- Analyze how perfect/imperfect competition between buyers and sellers of factors can impact wages, interest, and rents
- Compare the marginal productivity theory of income distribution versus real world income distribution
- Use the Lorenz Curve to analyze the distribution of income and wealth

297. Outcome: Factors of Production

What you'll learn to do: describe the incomes earned by the factors of production (land, labor, capital, entrepreneurship) wages, interest, rents, and profit

In this outcome, you will learn about income as it relates to both labor and capital markets.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: The Demand for Labor
- Case in Point: Computer Technology Increases the Demand for Some Workers and Reduces the Demand for Others
- Reading: The Supply of Labor
- Case in Point: An Airline Pilot's Lament
- Reading: Labor Markets at Work
- Case in Point: Technology and the Wage Gap
- Reading: Time and Interest Rates
- Reading: Interest Rates and Capital
- Reading: Loanable Funds
- Reading: Natural Resources and Conservation
- Self Check: Factors of Production

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

298. Reading: The Demand for Labor

Start Up: College Pays

On NBC's 2005 television series, *The Apprentice: Street Smarts vs. Book Smarts*, the contestants without college degrees who were chosen for the program were earning three times as much as those with college degrees. The two sides fought valiantly against each other, and the final episode pitted 37 year old, "street smart" Tana, a top-selling sales woman for Mary Kay, against 26 year old, "book smart" Kendra, a real estate agent. At the end of the Apprentice series, it was Kendra, the college graduate, to whom Donald Trump shouted, "You're hired!" As the array of contestants in the series demonstrates, not every college graduate earns more than every high school graduate, but on average, that is certainly the case.

One way of measuring the payoff from college is to compare the extent to which the wages of college-trained workers exceed the wages of high-school-trained workers. In the United States the payoff from college has soared over the last 25 years.

In 1979, male college graduates between 25 and 34 years old earned 28% more than male high school graduates in the same age bracket. By 2006 the gap had almost tripled—young male college graduates earned a stunning 76% more than young male high school graduates. Female college graduates gained as well. Young female college graduates earned 54% more than their high-school-educated counterparts in 1979; that gap increased to 86% by 2006.

The dramatic widening of the wage gap between workers with different levels of education reflects the operation of demand and supply in the market for labor. For reasons we will explore in this module, the demand for college graduates was increasing while the demand for high school graduates—particularly male high school graduates—was slumping.

Why would the demand curves for different kinds of labor shift? What determines the demand for labor? What about the supply? How do changes in demand and supply affect wages and employment? In this module we will apply what we have learned so far about production, profit maximization, and utility maximization to answer those questions in the context of a perfectly competitive market for labor.

This is the first of three modules focusing on factor markets, that is, on markets in which households supply factors of production—labor, capital, and natural resources—demanded by firms. Look back at the circular flow model introduced in the initial module on demand and supply. The bottom half of the circular flow model shows that households earn income from firms by supplying factors of production to them. The total income earned by households thus equals the total income earned by the labor, capital, and natural resources supplied to firms. Our focus in this module is on labor markets that operate in a competitive environment in which the individual buyers and sellers of labor are assumed to be price takers. Other modules on factor markets will discuss competitive markets for capital and for natural resources and imperfectly competitive markets for labor and for other factors of production.

Labor generates considerably more income in the economy than all other factors of production combined. Figure 12.1 shows the share of total income earned annually by workers in the United States since 1959. Labor accounts for roughly 73% of the income earned in the U.S. economy. The rest is generated by owners of capital and of natural resources.

Figure 12.1 *Labor's Share of U.S. Income, 1959–2007 . Workers have accounted for 70% of all the income earned in the United States since 1959. The remaining income was generated by capital and natural resources.*

We calculate the total income earned by workers by multiplying their average wage times the number of workers employed. We can view the labor market as a single market, as suggested in Panel (a) of Figure 12.2. Here we assume that all workers are identical, that there is a single market for them, and that they all earn the same wage, W; the level of employment is *L*. Although the assumption of a single labor market flies wildly in the face of reality, economists often use it to highlight broad trends in the market. For example, if we want to show the impact of an increase in the demand for labor throughout the economy, we can show labor as a single market in which the increase in demand raises wages and employment.

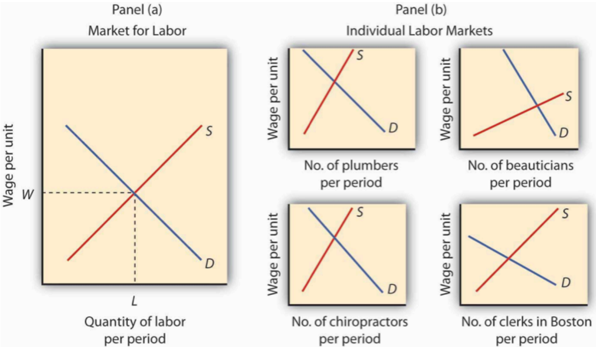


Figure 12.2 *Alternative Views of the Labor Market. One way to analyze the labor market is to assume that it is a single market with identical workers, as in Panel (a). Alternatively, we could examine specific pieces of the market, focusing on particular job categories or even on job categories in particular regions, as the graphs in Panel (b) suggest.*

But we can also use demand and supply analysis to focus on the market for a particular group of workers. We might examine the market for plumbers, beauticians, or chiropractors. We might even want to focus on the market for, say, clerical workers in the Boston area. In such cases, we would examine the demand for and the supply of a particular segment of workers, as suggested by the graphs in Panel (b) of Figure 12.2.

Macroeconomic analysis typically makes use of the highly aggregated approach to labor-market analysis illustrated in Panel (a), where labor is viewed as a single market. Microeconomic analysis typically assesses particular markets for labor, as suggested in Panel (b).

When we use the model of demand and supply to analyze the determination of wages and employment, we are assuming that market forces, not individuals, determine wages in the economy. The model says that equilibrium wages are determined by the intersection of the demand and supply curves for labor in a particular market. Workers and firms in the market are thus price takers; they take the market-determined wage as given and respond to it. We are, in this instance, assuming that perfect competition prevails in the labor market. Just as there are some situations in the analysis of markets for goods and services for which such an assumption is inappropriate, so there are some cases in which the assumption is inappropriate for labor markets. We examine such cases in a later module. In this module, however, we will find that the assumption of perfect competition can give us important insights into the forces that determine wages and employment levels for workers.

We will begin our analysis of labor markets in the next section by looking at the forces that influence the demand for labor. In the following section we will turn to supply. In the final section, we will use what we have learned to look at labor markets at work.

The Demand for Labor

A firm must have labor to produce goods and services. But how much labor will the firm employ? A profit-maximizing firm will base its decision to hire additional units of labor on the marginal decision rule: If the extra output that is produced by hiring one more unit of labor adds more to total revenue than it adds to total cost, the firm will increase profit by increasing its use of labor. It will continue to hire more and more labor up to the point that the extra revenue generated by the additional labor no longer exceeds the extra cost of the labor.

For example, if a computer software company could increase its annual total revenue by $50,000 by hiring a programmer at a cost of $49,000 per year, the marginal decision rule says that it should do so. Since the programmer will add $49,000 to total cost and $50,000 to total revenue, hiring the programmer will increase the company's profit by $1,000. If still another programmer would increase annual total revenue by $48,000 but would also add $49,000 to the firm's total cost, that programmer should not be hired because he or she would add less to total revenue than to total cost and would reduce profit.

Marginal Revenue Product and Marginal Factor Cost

The amount that an additional unit of a factor adds to a firm's total revenue during a period is called the **marginal revenue product (MRP)** of the factor. An additional unit of a factor of production adds to a firm's revenue in a two-step process: first, it increases the firm's output. Second, the increased output increases the firm's total revenue. We find marginal revenue product by multiplying the marginal product (MP) of the factor by the marginal revenue (MR).

$$MRP = MP \times MR$$

In a perfectly competitive market the marginal revenue a firm receives equals the market-determined price P. Therefore, for firms in perfect competition, we can express marginal revenue product as follows:

In perfect competition,

$$MRP = MP \times P$$

The marginal revenue product of labor (MRP_L) is the marginal product of labor (MP_L) times the marginal revenue (which is the same as price under perfect competition) the firm obtains from additional units of output that result from hiring the additional unit of labor. If an additional worker adds 4 units of output per day to a firm's production, and if each of those 4 units sells for $20, then the worker's marginal revenue product is $80 per day. With perfect competition, the marginal revenue product for labor, MRP_L, equals the marginal product of labor, MP_L, times the price, P, of the good or service the labor produces:

In perfect competition,

$$MRP_L = MP_L \times P$$

The law of diminishing marginal returns tells us that if the quantity of a factor is increased while other inputs are held constant, its marginal product will eventually decline. If marginal product is falling, marginal revenue product must be falling as well.

Suppose that an accountant, Stephanie Lancaster, has started an evening call-in tax advisory service. Between the hours of 7 p.m. and 10 p.m., customers can call and get advice on their income taxes. Ms. Lancaster's firm, TeleTax, is one of several firms offering similar advice; the going market price is $10 per call. Ms. Lancaster's business has expanded, so she hires other accountants to handle the calls. She must determine how many accountants to hire.

As Ms. Lancaster adds accountants, her service can take more calls. The table in Figure 12.3 gives the relationship between the number of accountants available to answer calls each evening and the number of calls TeleTax handles. Panel (a) shows the increase in the number of calls handled by each additional accountant—that accountant's marginal product. The first accountant can handle 13 calls per evening. Adding a second accountant increases the number of calls handled by 20. With two accountants, a degree of specialization is possible if each accountant takes calls dealing with questions about which he or she has particular expertise. Hiring the third accountant increases TeleTax's output per evening by 23 calls.

Suppose the accountants share a fixed facility for screening and routing calls. They also share a stock of reference materials to use in answering calls. As more accountants are added, the firm will begin to experience diminishing marginal returns. The fourth accountant increases output by 20 calls. The marginal product of additional accountants continues to decline after that. The marginal product curve shown in Panel (a) of Figure 12.3 thus rises and then falls.

Each call TeleTax handles increases the firm's revenues by $10. To obtain marginal revenue product, we multiply the marginal product of each accountant by $10; the marginal revenue product curve is shown in Panel (b) of Figure 12.3.

Labor	0	1	2	3	4	5	6	7	8	9	10	11	12
Output	0	13	33	56	76	93	106	116	123	128	131	133	134
Marginal product		13	20	23	20	17	13	10	7	5	3	2	1
Marginal revenue product		$130	200	230	200	170	130	100	70	50	30	20	10

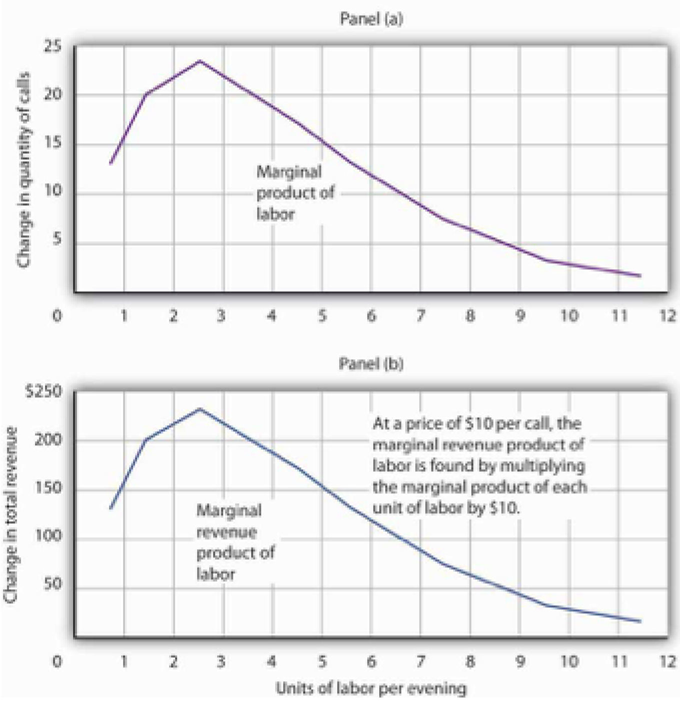


Figure 12.3. *Marginal Product and Marginal Revenue Product. The table gives the relationship between the number of accountants employed by TeleTax each evening and the total number of calls handled. From these values we derive the marginal product and marginal revenue product curves.*

We can use Ms. Lancaster's marginal revenue product curve to determine the quantity of labor she will hire. Suppose accountants in her area are available to offer tax advice for a nightly fee of $150. Each additional accountant Ms. Lancaster hires thus adds $150 per night to her total cost. The amount a factor adds to a firm's total cost per period is called its **marginal factor cost (MFC)**. Marginal factor cost (MFC) is the change in total cost (ΔTC) divided by the change in the quantity of the factor (Δf):

$$MFC = \frac{\Delta TC}{\Delta f}$$

The marginal factor cost to TeleTax of additional accountants ($150 per night) is shown as a horizontal line in Figure 12.4. It is simply the market wage (i.e., the price per unit of labor).

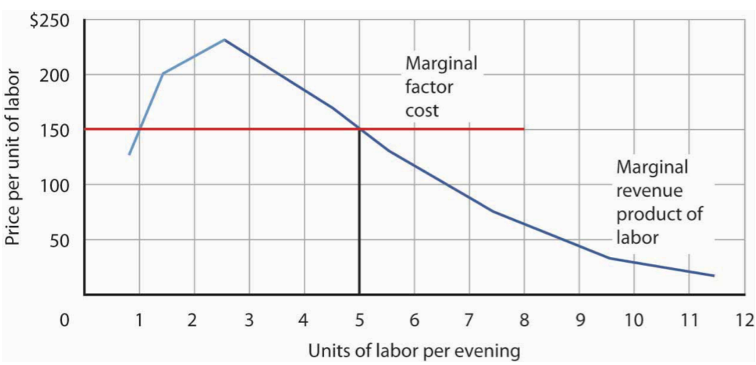


Figure 12.4. *Marginal Revenue Product and Demand. The downward-sloping portion of a firm's marginal revenue product curve is its demand curve for a variable factor. At a marginal factor cost of $150, TeleTax hires the services of five accountants.*

TeleTax will maximize profit by hiring additional units of labor up to the point where the downward-sloping portion of the marginal revenue product curve intersects the marginal factor cost curve; we see in Figure 12.4 that it will hire five accountants. Based on the information given in the table in Figure 12.3, we know that the five accountants will handle a total of 93 calls per evening; TeleTax will earn total revenue of $930 per evening. The firm pays $750 for the services of the five accountants—that leaves $180 to apply to the fixed cost associated with the tax advice service and the implicit cost of Stephanie Lancaster's effort in organizing the service. Recall that these implicit costs include the income forgone (that is, opportunity cost) by not shifting her resources, including her own labor, to her next best alternative.

If TeleTax had to pay a higher price for accountants, it would face a higher marginal factor cost curve and would hire fewer accountants. If the price were lower, TeleTax would hire more accountants. The downward-sloping portion of TeleTax's marginal revenue product curve shows the number of accountants it will hire at each price for accountants; it is thus the firm's demand curve for accountants. It is the portion of the curve that exhibits diminishing returns, and a firm will always seek to operate in the range of diminishing returns to the factors it uses.

It may seem counterintuitive that firms do not operate in the range of increasing returns, which would correspond to the upward-sloping portion of the marginal revenue product curve. However, to do so would forgo profit-enhancing opportunities. For example, in Figure 12.4, adding the second accountant adds $200 to revenue but only $150 to cost, so hiring that accountant clearly adds to profit. But why stop there? What about hiring a third accountant? That additional hire adds even more to revenue ($230) than to cost. In the region of increasing returns, marginal revenue product rises. With marginal factor cost constant, not to continue onto the downward-sloping part of the marginal revenue curve would be to miss out on profit-enhancing opportunities. The firm continues adding accountants until doing so no longer adds more to revenue than to cost, and that necessarily occurs where the marginal revenue product curve slopes downward.

In general, then, we can interpret the downward-sloping portion of a firm's marginal revenue product curve for a factor as its demand curve for that factor. Strictly speaking, it is only that part of the downward-sloping portion over which variable costs are at least covered. This is the flip-side of what you learned about a firm's supply curve in the module on competitive output markets: Only the portion of the rising marginal cost curve that lies above the minimum point of the average variable cost curve constitutes the supply curve of a perfectly competitive firm. We find the market demand for labor by adding the demand curves for individual firms.

The Two Rules Lead to the Same Outcome

In the module on competitive output markets we learned that profit-maximizing firms will increase output so long as doing so adds more to revenue than to cost, or up to the point where marginal revenue, which in perfect competition is the same as the market-determined price, equals marginal cost. In this module we have learned that profit-maximizing firms will hire labor up to the point where marginal revenue product equals marginal factor cost. Is it possible that a firm that follows the marginal decision rule for hiring labor would end up producing a different quantity of output compared to the quantity of output it would choose if it followed the marginal decision rule for deciding directly how much output to produce? Is there a conflict between these two marginal decision rules?

The answer is no. These two marginal decision rules are really just two ways of saying the same thing: one rule is in terms of quantity of output and the other in terms of the quantity of factors required to produce that quantity of output. Hiring an additional unit of a factor means producing a certain amount of additional output.

Using the example of TeleTax, at $150 per accountant per night, we found that Ms. Lancaster maximizes profit by hiring five accountants. The MP_L of the fifth accountant is ΔQ; it is 17. At five accountants, the marginal cost of a call is $\Delta TC/\Delta Q$ = $150/17 = $8.82, which is less than the price of $10 per call, so hiring that accountant adds to her profit. At six accountants, the marginal cost of a call would be $150/13 = $11.54, which is greater than the $10 price, so hiring a sixth accountant would lower profit. The profit-maximizing output of 93 calls, found by comparing marginal cost and price, is thus consistent with the profit-maximizing quantity of labor of five accountants, found by comparing marginal revenue product and marginal factor cost.

Shifts in Labor Demand

The fact that a firm's demand curve for labor is given by the downward-sloping portion of its marginal revenue product of labor curve provides a guide to the factors that will shift the curve. In perfect competition, marginal revenue product equals the marginal product of labor times the price of the good that the labor is involved in producing; anything that changes either of those two variables will shift the curve. The marginal revenue product of labor will change when there is a change in the quantities of other factors employed. It will also change as a result of a change in technology, a change in the price of the good being produced, or a change in the number of firms hiring the labor.

Changes in the Use of Other Factors of Production

As a firm changes the quantities of different factors of production it uses, the marginal product of labor may change. Having more reference manuals, for example, is likely to make additional accountants more productive—it will increase their marginal product. That increase in their marginal product would increase the demand for accountants. When an increase in the use of one factor of production increases the demand for another, the two factors are **complementary factors of production**.

One important complement of labor is human capital, the set of skills and abilities workers bring to the production of goods and services. When workers gain additional human capital, their marginal product rises. The demand for them by firms thus increases. This is perhaps one reason why you have decided to pursue a college education.

Other inputs may be regarded as substitutes for each other. A robot, for example, may substitute for some kinds of assembly-line labor. Two factors are **substitute factors of production** if the increased use of one lowers the demand for the other.

Changes in Technology

Technological changes can increase the demand for some workers and reduce the demand for others. The production of a more powerful computer chip, for example, may increase the demand for software engineers. It may also allow other production processes to be computerized and thus reduce the demand for workers who had been employed in those processes.

Technological changes have significantly increased the economy's output over the past century. The application of sophisticated technologies to production processes has boosted the marginal products of workers who have the skills these technologies require. That has increased the demand for skilled workers. The same technologies have been a substitute for less-skilled workers, and the demand for those workers has fallen. As the Case in Point on the impact of computer technology implies, envisioning the impact of technological change on demand for different kinds of labor may be something to keep in mind as you consider educational options. As you consider your major, for example, you should keep in mind that some occupations may benefit from technological changes; others may not.

Changes in Product Demand

A change in demand for a final product changes its price, at least in the short run. An increase in the demand for a product increases its price and increases the demand for factors that produce the product. A reduction in demand for a product reduces its price and reduces the demand for the factors used in producing it. Because the demand for factors that produce a product depends on the demand for the product itself, factor demand is said to be **derived demand**. That is, factor demand is derived from the demand for the product that uses the factor in its production.

Suppose, for example, that the demand for airplanes increases. The price and quantity of airplanes available will go up. A higher price for airplanes increases the marginal revenue product of labor of airplane-assembly workers and thus increases the demand for these workers.

Just as increases in the demand for particular goods or services increase the demand for the workers that produce them, so reductions in demand for particular goods or services will reduce the demand for the workers that produce them. An example is the relationship between the demand for train travel and the demand for conductors. Over the years, the fall in demand for train travel has reduced the demand for railroad conductors.

Changes in the Number of Firms

We can determine the demand curve for any factor by adding the demand for that factor by each of the firms using it. If more firms employ the factor, the demand curve shifts to the right. A reduction in the number of firms shifts the demand curve to the left. For example, if the number of restaurants in an area increases, the demand for waiters and waitresses in the area goes up. We expect to see local wages for these workers rise as a result.

299. Case in Point: Computer Technology Increases the Demand for Some Workers and Reduces the Demand for Others

Computer Technology Increases the Demand for Some Workers and Reduces the Demand for Others

"Moving an object, performing a calculation, communicating a piece of information or resolving a discrepancy. . . Which of these tasks can be performed by a computer?" ask economists David H. Autor, Frank Levy, and Richard J. Murname.

In general, computers are good at performing routine tasks and substitute for labor that had performed such tasks in the past. Conversely, computers are complements for workers performing nonroutine tasks, i.e., tasks that require such attributes as creativity, flexibility, and problem-solving. As the price of computers has fallen in recent decades, the demand for labor performing nonroutine tasks, usually college-educated workers, has grown, while the demand for labor performing routine tasks has fallen. The table below illustrates how computerization likely affects demand for different kinds of labor.

	Routine Tasks	Nonroutine Tasks
Analytic and Interactive Tasks		
Examples:	• Record-keeping • Calculation • Repetitive customer service (e.g. bank teller)	• Forming/testing hypotheses • Medical diagnosis • Legal writing • Persuading/selling • Managing others
Computer Impact:	• Substantial substitution	• Strong complementarities
Manual Task		
Examples:	• Picking or sording • Repetitive assembly	• Janitorial services • Truck driving
Computer Impact:	• Substantial substitution	• Limited opportunities for substitution or complementary

***Figure* 12.5.** *Predictions of Task Model for the Impact of Computerization on Four Categories of Workplace Tasks*

In studying the impact of computerization on labor demand, the study's authors have also noted that changes in the nature of certain tasks ("task-shifting") stemming from computerization have markedly changed what an occupation encompasses.

For example, the Department of Labor's Occupation Outlook Handbook in 1976 described what secretaries do as: "Secretaries relieve their employers of routine duties so they can work on more important matters. Although most secretaries type, take shorthand, and deal with callers, the time spent on these duties varies in different types of organizations." In contrast, the 2000 edition of the Handbook describes the work of secretaries quite differently: "As technology continues to expand in offices across the Nation, the role of the secretary has greatly evolved. Office automation and organizational restructuring have led secretaries to assume a wide range of new responsibilities once reserved for managerial and professional staff. Many secretaries now provide training and orientation to new staff, conduct research on the Internet, and learn to operate new office technologies." The authors find that this task-shifting within occupations, away from routine tasks and towards nonroutine tasks, is pervasive.

300. Reading: The Supply of Labor

The Supply of Labor

The demand for labor is one determinant of the equilibrium wage and equilibrium quantity of labor in a perfectly competitive market. The supply of labor, of course, is the other.

Economists think of the supply of labor as a problem in which individuals weigh the opportunity cost of various activities that can fill an available amount of time and choose how to allocate it. Everyone has 24 hours in a day. There are lots of uses to which we can put our time: we can raise children, work, sleep, play, or participate in volunteer efforts. To simplify our analysis, let us assume that there are two ways in which an individual can spend his or her time: in work or in leisure. Leisure is a type of consumption good; individuals gain utility directly from it. Work provides income that, in turn, can be used to purchase goods and services that generate utility.

The more work a person does, the greater his or her income, but the smaller the amount of leisure time available. An individual who chooses more leisure time will earn less income than would otherwise be possible. There is thus a tradeoff between leisure and the income that can be earned from work. We can think of the supply of labor as the flip side of the demand for leisure. The more leisure people demand, the less labor they supply.

Two aspects of the demand for leisure play a key role in understanding the supply of labor. First, leisure is a normal good. All other things unchanged, an increase in income will increase the demand for leisure. Second, the opportunity cost or "price" of leisure is the wage an individual can earn. A worker who can earn

$10 per hour gives up $10 in income by consuming an extra hour of leisure. The $10 wage is thus the price of an hour of leisure. A worker who can earn $20 an hour faces a higher price of leisure.

Income and Substitution Effects

Suppose wages rise. The higher wage increases the price of leisure. We saw in the module on consumer choice that consumers substitute more of other goods for a good whose price has risen. The substitution effect of a higher wage causes the consumer to substitute labor for leisure. To put it another way, the higher wage induces the individual to supply a greater quantity of labor.

We can see the logic of this substitution effect in terms of the marginal decision rule. Suppose an individual is considering a choice between extra leisure and the additional income from more work. Let MULe denote the marginal utility of an extra hour of leisure. What is the price of an extra hour of leisure? It is the wage W that the individual forgoes by not working for an hour. The extra utility of $1 worth of leisure is thus given by MULe/W.

Suppose, for example, that the marginal utility of an extra hour of leisure is 20 and the wage is $10 per hour. Then MULe/W equals 20/10, or 2. That means that the individual gains 2 units of utility by spending an additional $1 worth of time on leisure. For a person facing a wage of $10 per hour, $1 worth of leisure would be the equivalent of 6 minutes of leisure time.

Let MUY be the marginal utility of an additional $1 of income (Y is the abbreviation economists generally assign to income). The price of $1 of income is just $1, so the price of income PY is always $1. Utility is maximized by allocating time between work and leisure so that:

$$\frac{MU_Y}{P_Y} = \frac{MU_{\mathrm{Le}}}{W}$$

Now suppose the wage rises from W to W'. That reduces the marginal utility of $1 worth of leisure, MULe/W, so that the extra utility of earning $1 will now be greater than the extra utility of $1 worth of leisure:

$$\frac{MU_Y}{P_Y} > \frac{MU_{\text{Le}}}{W'}$$

Faced with the inequality above, an individual will give up some leisure time and spend more time working. As the individual does so, however, the marginal utility of the remaining leisure time rises and the marginal utility of the income earned will fall. The individual will continue to make the substitution until the two sides of the equation are again equal. For a worker, the substitution effect of a wage increase always reduces the amount of leisure time consumed and increases the amount of time spent working. A higher wage thus produces a positive substitution effect on labor supply.

But the higher wage also has an income effect. An increased wage means a higher income, and since leisure is a normal good, the quantity of leisure demanded will go up. And that means a reduction in the quantity of labor supplied.

For labor supply problems, then, the substitution effect is always positive; a higher wage induces a greater quantity of labor supplied. But the income effect is always negative; a higher wage implies a higher income, and a higher income implies a greater demand for leisure, and more leisure means a lower quantity of labor supplied. With the substitution and income effects working in opposite directions, it is not clear whether a wage increase will increase or decrease the quantity of labor supplied—or leave it unchanged.

Figure 12.6 illustrates the opposite pull of the substitution and income effects of a wage change facing an individual worker. A janitor, Meredith Wilson, earns $10 per hour. She now works 42 hours per week, on average, earning $420.

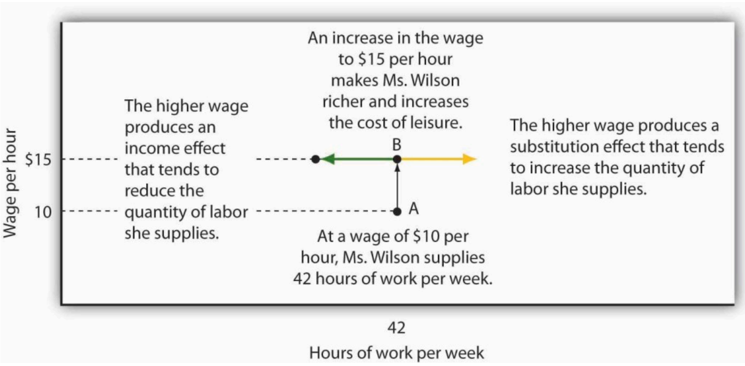


Figure 12.6. *The Substitution and Income Effects of a Wage Change. The substitution and income effects influence Meredith Wilson's supply of labor when she gets a pay raise. At a wage of $10 per hour, she supplies 42 hours of work per week (point A). At $15 per hour, the substitution effect pulls in the direction of an increased quantity of labor supplied, and the income effect pulls in the opposite direction.*

Now suppose Ms. Wilson receives a $5 raise to $15 per hour. As shown in Figure 12.6, the substitution effect of the wage change induces her to increase the quantity of labor she supplies; she substitutes some of her leisure time for additional hours of work. But she is richer now; she can afford more leisure. At a wage of $10 per hour, she was earning $420 per week. She could earn that same amount at the higher wage in just 28 hours. With her higher income, she can certainly afford more leisure time. The income effect of the wage change is thus negative; the quantity of labor supplied falls. The effect of the wage increase on the quantity of labor Ms. Wilson actually supplies depends on the relative strength of the substitution and income effects of the wage change. We will see what Ms. Wilson decides to do in the next section.

Wage Changes and the Slope of the Supply Curve

What would any one individual's supply curve for labor look like? One possibility is that over some range of labor hours supplied, the substitution effect will dominate. Because the marginal utility of leisure is relatively low when little labor is supplied (that is, when most time is devoted to leisure), it takes only a small increase in wages to induce the individual to substitute more labor for less leisure. Further, because few hours are worked, the income effect of those wage changes will be small.

Figure 12.7 shows Meredith Wilson's supply curve for labor. At a wage of $10 per hour, she supplies 42 hours of work per week (point A). An increase in her wage to $15 per hour boosts her quantity supplied to 48 hours per week (point B). The substitution effect thus dominates the income effect of a higher wage.

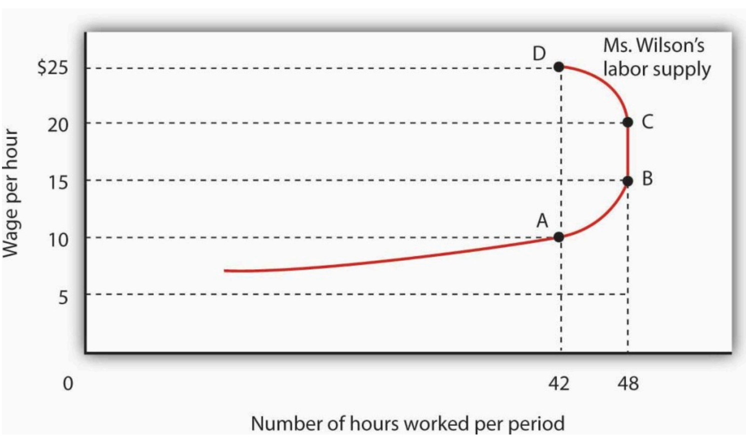


Figure 12.7. *A Backward-Bending Supply Curve for Labor. As the wage rate increases from $10 to $15 per hour, the quantity of labor Meredith Wilson supplies increases from 42 to 48 hours per week. Between points A and B, the positive substitution effect of the wage increase outweighs the negative income effect. As the wage rises above $15, the negative income effect just offsets the substitution effect, and Ms. Wilson's supply curve becomes a vertical line between points B and C. As the wage rises above $20, the income effect becomes stronger than the substitution effect, and the supply curve bends backward between points C and D.*

It is possible that beyond some wage rate, the negative income effect of a wage increase could just offset the positive substitution effect; over that range, a higher wage would have no effect on the quantity of labor supplied. That possibility is illustrated between points B and C on the supply curve in Figure 12.7; Ms. Wilson's supply curve is vertical. As wages continue to rise, the income effect becomes even stronger, and additional increases in the wage reduce the quantity of labor she supplies. The supply curve illustrated here bends backward beyond point C and thus assumes a negative slope. The supply curve for labor can thus slope upward over part of its range, become vertical, and then bend backward as the income effect of higher wages begins to dominate the substitution effect.

It is quite likely that some individuals have backward-bending supply curves for labor—beyond some point, a higher wage induces those individuals to work less, not more. However, supply curves for labor in specific labor markets are generally upward sloping. As wages in one industry rise relative to wages in other industries, workers shift their labor to the relatively high-wage one. An increased quantity of labor is supplied in that industry. While some exceptions have been found, the mobility of labor between competitive labor markets is likely to prevent the total number of hours worked from falling as the wage rate increases. Thus we shall assume that supply curves for labor in particular markets are upward sloping.

Shifts in Labor Supply

What events shift the supply curve for labor? People supply labor in order to increase their utility—just as they demand goods and services in order to increase their utility. The supply curve for labor will shift in response to changes in the same set of factors that shift demand curves for goods and services.

Changes in Preferences

A change in attitudes toward work and leisure can shift the supply curve for labor. If people decide they value leisure more highly, they will work fewer hours at each wage, and the supply curve for labor will shift to the left. If they decide they want more goods and services, the supply curve is likely to shift to the right.

Changes in Income

An increase in income will increase the demand for leisure, reducing the supply of labor. We must be careful here to distinguish movements along the supply curve from shifts of the supply curve itself. An income change resulting from a change in wages is shown by a movement along the curve; it produces the income and substitution effects we already discussed. But suppose income is from some other source: a person marries and has access to a spouse's income, or receives an inheritance, or wins a lottery. Those nonlabor increases in income are likely to reduce the supply of labor, thereby shifting the supply curve for labor of the recipients to the left.

Changes in the Prices of Related Goods and Services

Several goods and services are complements of labor. If the cost of child care (a complement to work effort) falls, for example, it becomes cheaper for workers to go to work, and the supply of labor tends to increase. If recreational activities (which are a substitute for work effort) become much cheaper, individuals might choose to consume more leisure time and supply less labor.

Changes in Population

An increase in population increases the supply of labor; a reduction lowers it. Labor organizations have generally opposed increases in immigration because their leaders fear that the increased number of workers will shift the supply curve for labor to the right and put downward pressure on wages.

Changes in Expectations

One change in expectations that could have an effect on labor supply is life expectancy. Another is confidence in the availability of Social Security. Suppose, for example, that people expect to live longer yet become less optimistic about their likely benefits from Social Security. That could induce an increase in labor supply.

Labor Supply in Specific Markets

The supply of labor in particular markets could be affected by changes in any of the variables we have already examined—changes in preferences, incomes, prices of related goods and services, population, and expectations. In addition to these variables that affect the supply of labor in general, there are changes that could affect supply in specific labor markets.

A change in wages in related occupations could affect supply in another. A sharp reduction in the wages of surgeons, for example, could induce more physicians to specialize in, say, family practice, increasing the supply of doctors in that field. Improved job opportunities for women in other fields appear to have decreased the supply of nurses, shifting the supply curve for nurses to the left.

The supply of labor in a particular market could also shift because of a change in entry requirements. Most states, for example, require barbers and beauticians to obtain training before entering the profession. Elimination of such requirements would increase the supply of these workers. Financial planners have, in recent years, sought the introduction of tougher licensing requirements, which would reduce the supply of financial planners.

Worker preferences regarding specific occupations can also affect labor supply. A reduction in willingness to take risks could lower the supply of labor available for risky occupations such as farm work (the most dangerous work in the United States), law enforcement, and fire fighting. An increased desire to work with children could raise the supply of child-care workers, elementary school teachers, and pediatricians.

KEY TAKEAWAYS

1. A higher wage increases the opportunity cost or price of leisure and increases worker incomes. The effects of these two changes pull the quantity of labor supplied in opposite directions.
2. A wage increase raises the quantity of labor supplied through the substitution effect, but it reduces the quantity supplied through the income effect. Thus an individual's supply curve of labor may be positively or negatively sloped, or have sections that are positively sloped, sections that are negatively sloped, and vertical sections. While some exceptions have been found, the labor supply curves for specific labor markets are generally upward sloping.

3. The supply curve for labor will shift as a result of a change in worker preferences, a change in nonlabor income, a change in the prices of related goods and services, a change in population, or a change in expectations.
4. In addition to the effects on labor supply of the variables just cited, other factors that can change the supply of labor in particular markets are changes in wages in related markets or changes in entry requirements.

301. Case in Point: An Airline Pilot's Lament

An Airline Pilot's Lament

Arguably, no single sector of the U.S. economy was hit harder by the events of 9/11 than the airline industry. By the time passengers did start returning, though, more than just the routine of getting through airport security had changed. Rather, the structure of the industry had begun to shift from domination by large national carriers—such as Delta, American, and United—operating according to the hub-and-spoke model, to increased competition from lower-cost regional carriers offering point-to-point service. Efforts by the large carriers in the early 2000s to reign in their costs and restore their financial health led to agreements with their labor unions that resulted in lower wages for most categories of airline workers.

How have airline employees responded to lower wages? Some categories of workers, such as mechanics, have little flexibility in deciding how many hours to work, but others, such as pilots, do. Below is an explanation by a female pilot who works for a major airline of the impact of wages cuts on her labor supply:

> We were normally scheduled for anywhere from 15 to 18 days a month, which translated into 80 to 95 hours of flying and around 280 hours of duty time. Duty time includes flight planning, preflighting, crew briefing, boarding, preflight checks of the airplane, etc. We bid for a monthly schedule that would fall somewhere in that range. After we were assigned our schedule for a month, we usually had the flexibility to drop or trade trips within certain constraints. Without going into the vast complexities of our contract, I

can tell you that, in general, we were allowed to drop down to 10 days a month, provided the company could cover the trips we wanted to drop, and still be considered a full-time employee. Generally, at that time, my goal was to work a minimum of 10 to 12 days a month and a maximum of 15 days a month. After the first round of pay cuts, the typical month became 16 to 20 days of flying. With that round of pay cuts, my general goal became to work a minimum of 15 days a month and a maximum of 17 days a month. I imagine that with another round of cuts my goal would be to keep my pay as high as I possibly can. Basically, I have a target income in mind. Anything above that was great, but I chose to have more days at home rather than more pay. As my target income became more difficult to achieve, I chose to work more days and hours to keep close to my target income. . . . When total compensation drops by more than 50% it is difficult to keep your financial head above water no matter how well you have budgeted.

302. Reading: Labor Markets at Work

Labor Markets at Work

We have seen that a firm's demand for labor depends on the marginal product of labor and the price of the good the firm produces. We add the demand curves of individual firms to obtain the market demand curve for labor. The supply curve for labor depends on variables such as population and worker preferences. Supply in a particular market depends on variables such as worker preferences, the skills and training a job requires, and wages available in alternative occupations. Wages are determined by the intersection of demand and supply. Once the wage in a particular market has been established, individual firms in perfect competition take it as given. Because each firm is a price taker, it faces a horizontal supply curve for labor at the market wage. For one firm, changing the quantity of labor it hires does not change the wage. In the context of the model of perfect competition, buyers and sellers are price takers. That means that a firm's choices in hiring labor do not affect the wage.

The operation of labor markets in perfect competition is illustrated in Figure 12.8. The wage W1 is determined by the intersection of demand and supply in Panel (a). Employment equals *L*1 units of labor per period. An individual firm takes that wage as given; it is the supply curve *s*1 facing the firm. This wage also equals the firm's marginal factor cost. The firm hires *l*1 units of labor, a quantity determined by the intersection of its marginal revenue product curve for labor *MRP*1 and the supply curve *s*1. We use lowercase letters to show quantity for a single firm and uppercase letters to show quantity in the market.

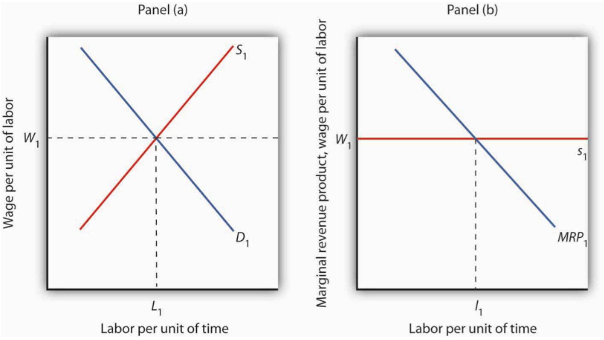


Figure 12.8*.* *Wage Determination and Employment in Perfect Competition. Wages in perfect competition are determined by the intersection of demand and supply in Panel (a). An individual firm takes the wage W1 as given. It faces a horizontal supply curve for labor at the market wage, as shown in Panel (b). This supply curve s1 is also the marginal factor cost curve for labor. The firm responds to the wage by employing l1 units of labor, a quantity determined by the intersection of its marginal revenue product*

curve MRP1 and its supply curve s1.

Changes in Demand and Supply

If wages are determined by demand and supply, then changes in demand and supply should affect wages. An increase in demand or a reduction in supply will raise wages; an increase in supply or a reduction in demand will lower them.

Panel (a) of Figure 12.9 shows how an increase in the demand for labor affects wages and employment. The shift in demand to *D*2 pushes the wage to *W*2and boosts employment to *L*2. Such an increase implies that the marginal product of labor has increased, that the number of firms has risen, or that the price of the good the labor produces has gone up. As we have seen, the marginal product of labor could rise because of an increase in the use of other factors of production, an improvement in technology, or an increase in human capital.

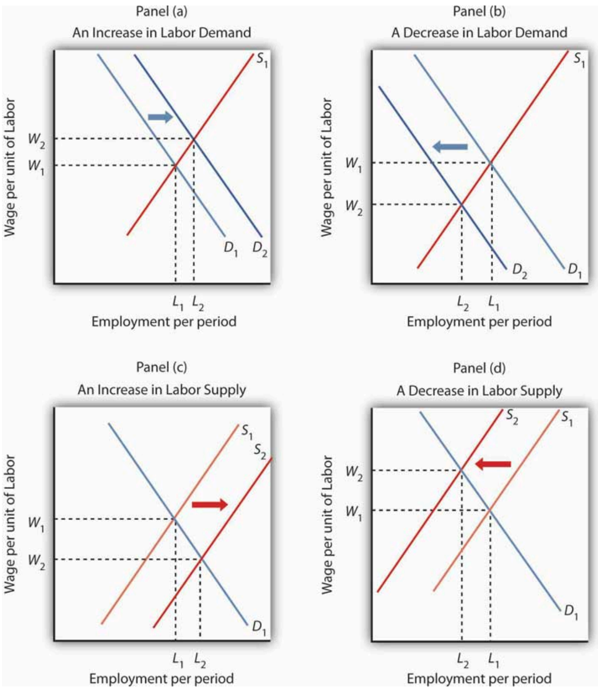


Figure 12.9. *Changes in the Demand for and Supply of Labor. Panel (a) shows an increase in demand for labor; the wage rises to W2 and employment rises to L2. A reduction in labor demand, shown in Panel (b), reduces employment and the wage level. An increase in the supply of labor, shown in Panel (c), reduces the wage to W2 and increases employment to L2. Panel (d) shows the effect of a reduction in the supply of labor; wages rise and employment falls.*

Clearly, a rising demand for labor has been the dominant trend in the market for U.S. labor through most of the nation's history. Wages and employment have generally risen as the availability of capital and other factors of production have increased, as technology has advanced, and as human capital has increased. All have increased the productivity of labor, and all have acted to increase wages.

Panel (b) of Figure 12.9 shows a reduction in the demand for labor to *D*2. Wages and employment both fall. Given that the demand for labor in the aggregate is generally increasing, reduced labor demand is most often found in specific labor markets. For example, a slump in construction activity in a particular community can lower the demand for construction workers. Technological changes can reduce as well as increase demand. The Case in Point on wages and technology suggests that technological changes since the late 1970s have tended to reduce the demand for workers with only a high school education while increasing the demand for those with college degrees.

Panel (c) of Figure 12.9 shows the impact of an increase in the supply of labor. The supply curve shifts to S2, pushing employment to *L*2 and cutting the wage to W2. For labor markets as a whole, such a supply increase could occur because of an increase in population or an increase in the amount of work people are willing to do. For individual labor markets, supply will increase as people move into a particular market.

Just as the demand for labor has increased throughout much of the history of the United States, so has the supply of labor. Population has risen both through immigration and through natural increases. Such increases tend, all other determinants of wages unchanged, to reduce wages. The fact that wages have tended to rise suggests that demand has, in general, increased more rapidly than supply. Still, the more supply rises, the smaller the increase in wages will be, even if demand is rising.

Finally, Panel (d) of Figure 12.9 shows the impact of a reduction in labor supply. One dramatic example of a drop in the labor supply was caused by a reduction in population after the outbreak of bubonic plague in Europe in 1348—the so-called Black Death. The plague killed about one-third of the people of Europe within a few years, shifting the supply curve for labor sharply to the left. Wages doubled during the period.

The fact that a reduction in the supply of labor tends to increase wages explains efforts by some employee groups to reduce labor supply. Members of certain professions have successfully promoted strict licensing requirements to limit the number of people who can enter the profession—U.S. physicians have been particularly successful in this effort. Unions often seek restrictions in immigration in an effort to reduce the supply of labor and thereby boost wages.

Competitive Labor Markets and the Minimum Wage

The Case in Point on technology and the wage gap points to an important social problem. Changes in technology boost the demand for highly educated workers. In turn, the resulting wage premium for more highly educated workers is a signal that encourages people to acquire more education. The market is an extremely powerful mechanism for moving resources to the areas of highest demand. At the same time, however, changes in technology seem to be leaving less educated workers behind. What will happen to people who lack the opportunity to develop the skills that the market values highly or who are unable to do so?

In order to raise wages of workers whose wages are relatively low, governments around the world have imposed minimum wages. A minimum wage works like other price floors. The impact of a minimum wage is shown in Panel (a) of Figure 12.10. Suppose the current equilibrium wage of unskilled workers is W1, determined by the intersection of the demand and supply curves of these workers. The government determines that this wage is too low and orders that it be increased to Wm, a minimum wage. This strategy reduces employment from *L*1 to *L*2, but it raises the incomes of those who continue to work. The higher wage also increases the quantity of labor supplied to *L*3. The gap between the quantity of labor supplied and the quantity demanded, *L*3− *L*2, is a surplus—a surplus that increases unemployment.

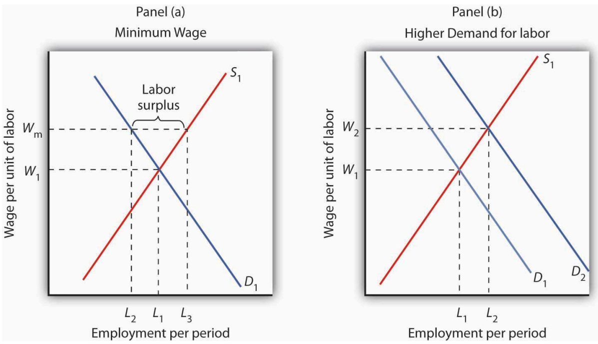


Figure 12.10. *Alternative Responses to Low Wages. Government can respond to a low wage by imposing a minimum wage of Wm in Panel (a). This increases the quantity of labor supplied and reduces the quantity demanded. It does, however, increase the income of those who keep their jobs. Another way the government can boost wages is by raising the demand for labor in Panel (b). Both wages and employment rise.*

Some economists oppose increases in the minimum wage on grounds that such increases boost unemployment. Other economists argue that the demand for unskilled labor is relatively inelastic, so a higher minimum wage boosts the incomes of unskilled workers as a group. That gain, they say, justifies the policy, even if it increases unemployment.

An alternative approach to imposing a legal minimum is to try to boost the demand for labor. Such an approach is illustrated in Panel (b). An increase in demand to $D2$ pushes the wage to $W2$ and at the same time increases employment to $L2$. Public sector training programs that seek to increase human capital are examples of this policy.

Still another alternative is to subsidize the wages of workers whose incomes fall below a certain level. Providing government subsidies—either to employers who agree to hire unskilled workers or to workers themselves in the form of transfer payments—enables people who lack the skills—and the ability to acquire the skills—needed to earn a higher wage to earn more without the loss of jobs implied by a higher minimum wage. Such programs can be costly. They also reduce the incentive for low-skilled workers to develop the skills that are in greater demand in the marketplace.

KEY TAKEAWAYS

1. Wages in a competitive market are determined by demand and supply.
2. An increase in demand or a reduction in supply will increase the equilibrium wage. A reduction in demand or an increase in supply will reduce the equilibrium wage.
3. The government may respond to low wages for some workers by imposing the minimum wage, by attempting to increase the demand for those workers, or by subsidizing the wages of workers whose incomes fall below a certain level.

303. Case in Point: Technology and the Wage Gap

Technology and the Wage Gap

Economist Daron Acemoglu's research begins by noting that the college premium, defined as the average wages of college graduates relative to that of high school graduates, rose 25% between 1979 and 1995. Also, during essentially the same period, wage inequality rose. Whereas in the early 1970s, a person in the 90th percentile of the wage distribution earned 266% more than a person in the 10th percentile earned, 25 years later the gap had increased to 366%. The consensus view maintains that the increase in the college premium and in wage inequality stem primarily from skill-biased technological change. Skill-biased technological change means that, in general, newly developed technologies have favored the hiring of workers with better education and more skills.

But, while technological advances *may* increase the demand for skilled workers, the opposite can also occur. For example, the rise of factories, assembly lines, and interchangeable parts in the 19th century reduced the demand for skilled artisans such as weavers and watchmakers. So, the 20th century skill-bias of technological change leads researchers to ask why recent technological change has taken the form it has.

Acemoglu's answer is that, at least in part, the character of technological change itself constitutes a response to profit incentives:

> *The early nineteenth century was characterized by skill-*

> *replacing developments because the increased supply of unskilled workers in the English cities (resulting from migration from rural areas and from Ireland) made the introduction of these technologies profitable. In contrast, the twentieth century has been characterized by skill-biased technical change because the rapid increase in the supply of skilled workers has induced the development of skill-complementary technologies. (p. 9)*

In general, technological change in this model is *endogenous*—that is, its character is shaped by any incentives that firms face.

Of course, an increase in the supply of skilled labor, as has been occurring relentlessly in the U.S. over the past century, would, other things unchanged, lead to a fall in the wage premium. Acemoglu and others argue that the increase in the demand for skilled labor has simply outpaced the increase in supply.

But this also begs the *why* question. Acemoglu's answer again relies on the profit motive:

> *The development of skill-biased technologies will be more profitable when they have a larger market size—i.e., when there are more skilled workers. Therefore, the equilibrium degree of skill bias could be an increasing function of the relative supply of skilled workers. An increase in the supply of skills will then lead to skill-biased technological change. Furthermore, acceleration in the supply of skills can lead to acceleration in the demand for skill." (p. 37).*

It follows from this line of reasoning that the rapid increase in the supply of college-educated workers led to more skill-biased technologies that in turn led to a higher college premium.

While the above ideas explain the college premium, they do not address why the real wages of low-skilled workers have fallen in recent decades. Popular explanations include the decreased role of labor unions and the increased role of international trade. Many studies, though, have concluded that the direct impacts of these factors have been limited. For example, in both the U.S. and U.K., rising wage inequality preceded the decline of labor unions. Concerning the impact of trade on inequality, economist John Bound observed, "The wage gap widened in a broad range of industries, including the service sector, and that cannot be explained by a shift in international trade…For example, the gap between the wages of high school–educated and college-educated workers widened in hospitals, and they aren't affected by foreign production." While Acemoglu accepts those conclusions, he argues that labor market institutions and trade may have interacted with technological change to magnify technological change's direct effect on inequality. For example, skill-biased technological change makes wage compression that unions tend to advocate more costly for skilled workers and thus weakens the "coalition between skilled and unskilled work that maintains unions" (p. 52). Likewise, trade expansion with less developed countries may have led to more skill-biased technological change than otherwise would have occurred.

Acemoglu recognizes that more research is needed to determine whether these indirect effects are operating and, if they are, the sizes of these effects, but looking at how technological change responds to economic conditions may begin to solve some heretofore puzzling aspects of recent labor market changes.

304. Reading: Time and Interest Rates

Start Up: Building the "Internet in the Sky"

The race to build the "Internet in the Sky" started in the early 1990s. One plan was to build 840 low earth-orbiting (LEO) satellites that would allow information to be sent and received instantaneously anywhere on the face of the globe. At least that was the plan.

A number of telecommunication industry giants, as well as some large manufacturing companies, were impressed with the possibilities. They saw what they thought was a profitable opportunity and decided to put up some financial capital. Craig McCaw, who made a fortune developing and then selling to AT&T, the world's largest cellular phone network, became chair of Teledesic, the company he formed to build the LEO satellite system. McCaw put up millions of dollars to fund the project, as did Microsoft's Bill Gates and Prince Alwaleed Bin Talal Bin Abdulaziz of Saudi Arabia. Boeing, Motorola, and Matra Marconi Space, Europe's leading satellite manufacturer, became corporate partners. Altogether, the company raised almost a billion dollars. The entire project was estimated to cost $9 billion.

But, alas, a decade later the company had shifted into very low gear. From the initial plan for 840 satellites, the project was scaled back to 300 satellites and then to a mere 30. Then, in 2003 in a letter to the U.S. Federal Communications commission, it announced that it was giving up its license to use a large part of the radio spectrum.

What happened to this dream? The development of cellular networks to handle data and video transmissions may have made the satellite system seem unnecessary. In contrast to a satellite system that has to be built in total in order to bring in a single customer, wireless companies were able to build their customer base city by city.

Even if the project had become successful, the rewards to the companies and to the individuals that put their financial capital into the venture would have been a long time in coming. Service was initially scheduled to begin in 2001, but Teledesic did not even sign a contract to build its first two satellites until February 2002, and six months later the company announced that work on those had been suspended.

Teledesic's proposed venture was bigger than most capital projects, but it shares some basic characteristics with any acquisition of capital by firms. The production of capital—the goods used in producing other goods and services—requires sacrificing consumption. The returns to capital will be spread over the period in which the capital is used. The choice to acquire capital is thus a choice to give up consumption today in hopes of returns in the future. Because those returns are far from certain, the choice to acquire capital is inevitably a risky one.

For all its special characteristics, however, capital is a factor of production. As we investigate the market for capital, the concepts of marginal revenue product, marginal factor cost, and the marginal decision rule that we have developed will continue to serve us. The big difference is that the benefits and costs of holding capital are distributed over time.

We will also examine markets for natural resources in this module. Like decisions involving capital, choices in the allocation of natural resources have lasting effects. For potentially exhaustible natural resources such as oil, the effects of those choices last forever.

For the analysis of capital and natural resources, we shift from the examination of outcomes in the current period to the analysis of outcomes distributed over many periods. Interest rates, which link the values of payments that occur at different times, will be central to our analysis.

Time and Interest Rates

Time, the saying goes, is nature's way of keeping everything from happening all at once. And the fact that everything does not happen at once introduces an important complication in economic analysis.

When a company decides to use funds to install capital that will not begin to produce income for several years, it needs a way to compare the significance of funds spent now to income earned later. It must find a way to compensate financial investors who give up the use of their funds for several years, until the project begins to pay off. How can payments that are distributed across time be linked to one another? Interest rates are the linkage mechanism; we shall investigate how they achieve that linkage in this section.

The Nature of Interest Rates

Consider a delightful problem of choice. Your Aunt Carmen offers to give you $10,000 now or $10,000 in one year. Which would you pick?

Most people would choose to take the payment now. One reason for that choice is that the average level of prices is likely to rise over the next year. The purchasing power of $10,000 today is thus greater than the purchasing power of $10,000 a year hence. There is also a question of whether you can count on receiving the payment. If you take it now, you have it. It is risky to wait a year; who knows what will happen?

Let us eliminate both of these problems. Suppose that you are confident that the average level of prices will not change during the year, and you are absolutely certain that if you choose to wait for the payment, you and it will both be available. Will you take the payment now or wait?

Chances are you would *still* want to take the payment now. Perhaps there are some things you would like to purchase with it, and you would like them sooner rather than later. Moreover, if you wait a year to get the payment, you will not be able to use it while you are waiting. If you take it now, you can choose to spend it now or wait.

Now suppose Aunt Carmen wants to induce you to wait and changes the terms of her gift. She offers you $10,000 now or $11,000 in one year. In effect, she is offering you a $1,000 bonus if you will wait a year. If you agree to wait a year to receive Aunt Carmen's payment, you will be accepting her *promise* to provide funds instead of the funds themselves. Either will increase your wealth, which is the sum of all your assets less all your liabilities. Assets are anything you have that is of value; liabilities are obligations to make future payments. Both a $10,000 payment from Aunt Carmen now and her promise of $11,000 in a year are examples of assets. The alternative to holding wealth is to consume it. You could, for example, take Aunt Carmen's $10,000 and spend it for a trip to Europe, thus reducing your wealth. By making a better offer—$11,000 instead of $10,000—Aunt Carmen is trying to induce you to accept an asset you will not be able to consume during the year.

The $1,000 bonus Aunt Carmen is offering if you will wait a year for her payment is interest. In general, interest is a payment made to people who agree to postpone their use of wealth. The interest rate represents the opportunity cost of using wealth today, expressed as a percentage of the amount of wealth whose use is postponed. Aunt Carmen is offering you $1,000 if you will pass up the $10,000 today. She is thus offering you an interest rate of 10% ($1,000/$10,000 = 0.1 = 10%).

Suppose you tell Aunt Carmen that, given the two options, you would still rather have the $10,000 today. She now offers you $11,500 if you will wait a year for the payment—an interest rate of 15% ($1,500/$10,000=0.15=15%). The more she pays for waiting, the higher the interest rate.

You are probably familiar with the role of interest rates in loans. In a loan, the borrower obtains a payment now in exchange for promising to repay the loan in the future. The lender thus must postpone his or her use of wealth until the time of repayment. To induce lenders to postpone their use of their wealth, borrowers offer interest. Borrowers are willing to pay interest because it allows them to acquire the sum now rather than having to wait for it. And lenders require interest payments to compensate them for postponing their own use of their wealth.

Interest Rates and Present Value

People generally prefer to receive a payment of some amount today rather than wait to receive that same amount later. We may conclude that the value today of a payment in the future is less than the dollar value of the future payment. An important application of interest rates is to show the relationship between the current and future values of a particular payment.

To see how we can calculate the current value of a future payment, let us consider an example similar to Aunt Carmen's offer. This time you have $1,000 and you deposit it in a bank, where it earns interest at the rate of 10% per year.

How much will you have in your bank account at the end of one year? You will have the original $1,000 plus 10% of $1,000, or $1,100: $1,000 + (0.10)($1,000) = $1,100

More generally, if we let P_0 equal the amount you deposit today, r the percentage rate of interest, and P_1 the balance of your deposit at the end of 1 year, then we can write:

$$P_0 + rP_0 = P_1$$

Factoring out the P_0 term on the left-hand side of the equation above, we have:

$$P_0(1+r) = P_1$$

This equation shows how to determine the future value of a payment or deposit made today. Now let us turn the question around. We can ask what P_1, an amount that will be available 1 year from now, is worth today. We solve for this by dividing both sides by $(1 + r)$ to obtain:

$$P_0 = \frac{P_1}{(1+r)}$$

This suggests how we can compute the value today, P_0, of an amount P_1 that will be paid a year hence. An amount that would equal a particular future value if deposited today at a specific interest rate is called the present value of that future value.

More generally, the present value of any payment to be received n periods from now equals

$$P_0 = \frac{P_n}{(1+r)^n}$$

Suppose, for example, that your Aunt Carmen offers you the option of \$1,000 now or \$15,000 in 30 years. We can use this equation to help you decide which sum to take. The present value of \$15,000 to be received in 30 years, assuming an interest rate of 10%, is:

$$P_0 = \frac{P_{30}}{(1+r)^{30}} = \frac{\$15,000}{(1+0.10)^{30}} = \$859.63$$

Assuming that you could earn that 10% return with certainty, you would be better off taking Aunt Carmen's \$1,000 now; it is greater than the present value, at an interest rate of 10%, of the \$15,000 she would give you in 30 years. The \$1,000 she gives you now, assuming an interest rate of 10%, in 30 years will grow to:

$$\$1,000(1 + 0.10)^{30} = \$17,449.40$$

The present value of some future payment depends on three things.

1. **The Size of the Payment Itself.**The bigger the future payment, the greater its present value.
2. **The Length of the Period Until Payment.**The present value

depends on how long a period will elapse before the payment is made. The present value of $15,000 in 30 years, at an interest rate of 10%, is $859.63. But that same sum, if paid in 20 years, has a present value of $2,229.65. And if paid in 10 years, its present value is more than twice as great: $5,783.15. The longer the time period before a payment is to be made, the lower its present value.

3. **The Rate of Interest.**The present value of a payment of $15,000 to be made in 20 years is $2,229.65 if the interest rate is 10%; it rises to $5,653.34 at an interest rate of 5%. The lower the interest rate, the higher the present value of a future payment. Table 13.1 gives present values of a payment of $15,000 at various interest rates and for various time periods.

Table 13.1 Time, Interest Rates, and Present Value

	Present Value of $15,000			
Interest rate (%)	**Time until payment**			
	5 years	**10 years**	**15 years**	**20 years**
5	$11,752.89	$9,208.70	$7,215.26	$5,653.34
10	9,313.82	5,783.15	3,590.88	2,229.65
15	7,457.65	3,707.77	1,843.42	916.50
20	6,028.16	2,422.58	973.58	391.26

The higher the interest rate and the longer the time until payment is made, the lower the present value of a future payment. The table below shows the present value of a future payment of $15,000 under different conditions. The present value of $15,000 to be paid in five years is $11,752.89 if the interest rate is 5%. Its present value is just $391.26 if it is to be paid in 20 years and the interest rate is 20%.

The concept of present value can also be applied to a series of future payments. Suppose you have been promised $1,000 at the end of each of the next 5 years. Because each payment will occur at a different time, we calculate the present value of the series of payments by taking the value of each payment separately and adding them together. At an interest rate of 10%, the present value P_0 is:

$$P_0 = \frac{\$1,000}{1.10} + \frac{\$1,000}{(1.10)^2} + \frac{\$1,000}{(1.10)^3} + \frac{\$1,000}{(1.10)^4} + \frac{\$1,000}{(110)^5} = \$3,790.78$$

Interest rates can thus be used to compare the values of payments that will occur at different times. Choices concerning capital and natural resources require such comparisons, so you will find applications of the concept of present value throughout this module, but the concept of present value applies whenever costs and benefits do not all take place in the current period.

State lottery winners often have a choice between a single large payment now or smaller payments paid out over a 25- or 30-year period. Comparing the single payment now to the present value of the future payments allows winners to make informed decisions. For example, in June 2005 Brad Duke, of Boise, Idaho, became the winner of one of the largest lottery prizes ever. Given the alternative of claiming the $220.3 million jackpot in 30 annual payments of $7.4 million or taking $125.3 million in a lump sum, he chose the latter. Holding unchanged all other considerations that must have been going through his mind, he must have thought his best rate of return would be greater than 4.17%. Why 4.17%? Using an interest rate of 4.17%, $125.3 million is equal to slightly less than the present value of the 30-year stream of payments. At all interest rates greater than 4.17%, the present value of the stream of benefits would be less than $125.3 million. At all interest rates less than 4.17%, the present value of the stream of payments would be more than $125.3 million. Our present value analysis suggests that if he thought the interest rate he could earn was more than 4.17%, he should take the lump sum payment, which he did.

Case in Point: Waiting for Death and Life Insurance

It is a tale that has become all too familiar.

Call him Roger Johnson. He has just learned that his cancer is not treatable and that he has only a year or two to live. Mr. Johnson is unable to work, and his financial burdens compound his tragic medical situation. He has mortgaged his house and sold his other assets in a desperate effort to get his hands on the cash he needs for care, for food, and for shelter. He has a life insurance policy, but it will pay off only when he dies. If only he could get some of that money sooner…

The problem facing Mr. Johnson has spawned a market solution—companies and individuals that buy the life insurance policies of the terminally ill. Mr. Johnson could sell his policy to one of these companies or individuals and collect the purchase price. The buyer takes over his premium payments. When he dies, the company will collect the proceeds of the policy.

The industry is called the viatical industry (the term *viatical* comes from *viaticum*, a Christian sacrament given to a dying person). It provides the terminally ill with access to money while they are alive; it provides financial investors a healthy interest premium on their funds.

It is a chilling business. Potential buyers pore over patient's medical histories, studying T-cell counts and other indicators of a patient's health. From the buyer's point of view, a speedy death is desirable, because it means the investor will collect quickly on the purchase of a patient's policy.

A patient with a life expectancy of less than six months might be able to sell his or her life insurance policy for 80% of the face value. A $200,000 policy would thus sell for $160,000. A person with a better prognosis will collect less. Patients expected to live two years, for example, might get only 60% of the face value of their policies.

Are investors profiting from the misery of others? Of course they are. But, suppose that investors refused to take advantage of the misfortune of the terminally ill. That would deny dying people the chance to acquire funds that they desperately need. As is the case with all voluntary exchange, the viatical market creates win-win situations. Investors "win" by earning high rates of return on their investment. And the dying patient? He or she is in a terrible situation, but the opportunity to obtain funds makes that person a "winner" as well.

Kim D. Orr, a former agent with Life Partners Inc. (www.lifepartnersinc.com), one of the leading firms in the viatical industry, recalled a case in his own family. "Some years ago, I had a cousin who died of AIDS. He was, at the end, destitute and had to rely totally on his family for support. Today, there is a broad market with lots of participants, and a patient can realize a high fraction of the face value of a policy on selling it. The market helps buyers and patients alike."

In recent years, this industry has been renamed the life settlements industry, with policy transfers being offered to healthier, often elderly, policyholders. These healthier individuals are sometimes turning over their policies for a payment to third parties who pay the premiums and then collect the benefit when the policyholders die. Expansion of this practice has begun to raise costs for life insurers, who assumed that individuals would sometimes let their policies lapse, with the result that the insurance company does not have to pay claims on them. Businesses buying life insurance policies are not likely to let them lapse.

305. Reading: Interest Rates and Capital

The Demand for Capital

The quantity of capital that firms employ in their production of goods and services has enormously important implications for economic activity and for the standard of living people in the economy enjoy. Increases in capital increase the marginal product of labor and boost wages at the same time they boost total output. An increase in the stock of capital therefore tends to raise incomes and improve the standard of living in the economy.

Capital is often a fixed factor of production in the short run. A firm cannot quickly retool an assembly line or add a new office building. Determining the quantity of capital a firm will use is likely to involve long-run choices.

A firm uses additional units of a factor until marginal revenue product equals marginal factor cost. Capital is no different from other factors of production, save for the fact that the revenues and costs it generates are distributed over time. As the first step in assessing a firm's demand for capital, we determine the present value of marginal revenue products and marginal factor costs.

Capital and Net Present Value

Suppose Carol Stein is considering the purchase of a new \$95,000 tractor for her farm. Ms. Stein expects to use the tractor for five years and then sell it; she expects that it will sell for \$22,000 at the end of the five-year period. She has the \$95,000 on hand now; her alternative to purchasing the tractor could be to put $95,000 in a bond account earning 7% annual interest.

Ms. Stein expects that the tractor will bring in additional annual revenue of $50,000 but will cost $30,000 per year to operate, for net revenue of $20,000 annually. For simplicity, we shall suppose that this net revenue accrues at the end of each year.

Should she buy the tractor? We can answer this question by computing the tractor's net present value (NPV), which is equal to the present value of all the revenues expected from an asset minus the present value of all the costs associated with it. We thus measure the difference between the present value of marginal revenue products and the present value of marginal factor costs. If NPV is greater than zero, purchase of the asset will increase the profitability of the firm. A negative NPV implies that the funds for the asset would yield a higher return if used to purchase an interest-bearing asset. A firm will maximize profits by acquiring additional capital up to the point that the present value of capital's marginal revenue product equals the present value of marginal factor cost.

If the revenues generated by an asset in period *n* equal Rn and the costs in period *n* equal Cn, then the net present value NPV0 of an asset expected to last for *n* years is:

$$NPV_0 = R_0 - C_0 + \frac{R_1 - C_1}{1 + r} + \dots + \frac{R_n - C_n}{(1 + r)^n}$$

To purchase the tractor, Ms. Stein pays $95,000. She will receive additional revenues of $50,000 per year from increased planting and more efficient harvesting, less the operating cost per year of $30,000, plus the $22,000 she expects to get by selling the tractor at the end of five years. The net present value of the tractor, NPV0 is thus given by:

$$NPV_0 = -\$95{,}000 + \frac{\$20{,}000}{1.07} + \frac{\$20{,}000}{1.07^2} + \frac{\$20{,}000}{1.07^3} + \frac{\$20{,}000}{1.07^4} + \frac{\$42{,}000}{1.07^5} = \$2{,}690$$

Given the cost of the tractor, the net returns Ms. Stein projects, and an interest rate of 7%, Ms. Stein will increase her profits by purchasing the tractor. The tractor will yield a return whose present value is $2,690 greater than the return that could be obtained by the alternative of putting the $95,000 in a bond account yielding 7%.

Ms. Stein's acquisition of the tractor is called investment. Economists define investment as an addition to capital stock. Any acquisition of new capital goods therefore qualifies as investment.

The Demand Curve for Capital

Our analysis of Carol Stein's decision regarding the purchase of a new tractor suggests the forces at work in determining the economy's demand for capital. In deciding to purchase the tractor, Ms. Stein considered the price she would have to pay to obtain the tractor, the costs of operating it, the marginal revenue product she would receive by owning it, and the price she could get by selling the tractor when she expects to be done with it. Notice that with the exception of the purchase price of the tractor, *all* those figures were projections. Her decision to purchase the tractor depends almost entirely on the costs and benefits she *expects* will be associated with its use.

Finally, Ms. Stein converted all those figures to a net present value based on the interest rate prevailing at the time she made her choice. A positive NPV means that her profits will be increased by purchasing the tractor. That result, of course, depends on the prevailing interest rate. At an interest rate of 7%, the NPV is positive.

At an interest rate of 8%, the NPV would be negative. At that interest rate, Ms. Stein would do better to put her funds elsewhere.

At any one time, millions of choices like that of Ms. Stein concerning the acquisition of capital will be under consideration. Each decision will hinge on the price of a particular piece of capital, the expected cost of its use, its expected marginal revenue product, its expected scrap value, and the interest rate. Not only will firms be considering the acquisition of new capital, they will be considering retaining existing capital as well. Ms. Stein, for example, may have other tractors. Should she continue to use them, or should she sell them? If she keeps them, she will experience a stream of revenues and costs over the next several periods; if she sells them, she will have funds now that she could use for something else. To decide whether a firm should keep the capital it already has, we need an estimate of the NPV of each unit of capital. Such decisions are always affected by the interest rate. At higher rates of interest, it makes sense to sell some capital rather than hold it. At lower rates of interest, the NPV of holding capital will rise.

Because firms' choices to acquire new capital and to hold existing capital depend on the interest rate, the demand curve for capital in Figure 13.1, which shows the quantity of capital firms intend to hold at each interest rate, is downward-sloping. At point A, we see that at an interest rate of 10%, $8 trillion worth of capital is demanded in the economy. At point B, a reduction in the interest rate to 7% increases the quantity of capital demanded to $9 trillion. At point C, at an interest rate of 4%, the quantity of capital demanded is $10 trillion. A reduction in the interest rate increases the quantity of capital demanded.

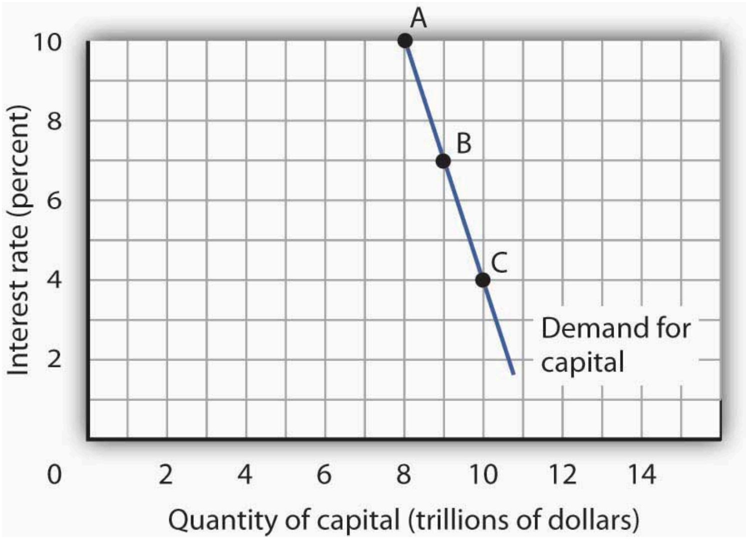


Figure 13.1. *The Demand Curve for Capital. The quantity of capital firms will want to hold depends on the interest rate. The higher the interest rate, the less capital firms will want to hold.*

The demand curve for capital for the economy is found by summing the demand curves of all holders of capital. Ms. Stein's demand curve, for example, might show that at an interest rate of 8%, she will demand the capital she already has—suppose it is $600,000 worth of equipment. If the interest rate drops to 7%, she will add the tractor; the quantity of capital she demands rises to $695,000. At interest rates greater than 8%, she might decide to reduce her maintenance efforts for some of the capital she already has; the quantity of capital she demands would fall below $600,000. As with the demand for capital in the economy, we can expect individual firms to demand a smaller quantity of capital when the interest rate is higher.

Shifts in the Demand for Capital

Why might the demand for capital change? Because the demand for capital reflects the marginal revenue product of capital, anything that changes the marginal revenue product of capital will shift the demand for capital. Our search for demand shifters must thus focus on factors that change the marginal product of capital, the prices of the goods capital produces, and the costs of acquiring and holding capital. Let us discuss some factors that could affect these variables and thus shift the demand for capital.

Changes in Expectations

Choices concerning capital are always based on expectations. Net present value is computed from the expected revenues and costs over the expected life of an asset. If firms' expectations change, their demand for capital will change. If something causes firms to revise their sales expectations upward (such as stronger than expected sales in the recent past), it is likely to increase their demand for capital. Similarly, an event that dampens firms' expectations (such as recent weak sales) is likely to reduce their demand for capital.

Technological Change

Technological changes can increase the marginal product of capital and thus boost the demand for capital. The discovery of new ways to integrate computers into production processes, for example, has dramatically increased the demand for capital in the last few years. Many universities are adding new classroom buildings or renovating old ones so they can better use computers in instruction, and businesses use computers in nearly every facet of operations.

Changing Demand for Goods and Services

Ultimately, the source of demand for factors of production is the demand for the goods and services produced by those factors. Economists say that the demand for a factor is a "derived" demand—derived, that is, from the demand for what the factor produces. As population and incomes expand, we can expect greater demand for goods and services, a change that will increase the demand for capital.

Changes in Relative Factor Prices

Firms achieve the greatest possible output for a given total cost by operating where the ratios of marginal product to factor price are equal for all factors of production. For a firm that uses labor (L) and capital (K), for example, this requires that $MPL/PL=MPK/PK$, where $MPL$and MPK are the marginal products of labor and capital, respectively, and PL and PK are the prices of labor and capital, respectively. Suppose these equalities hold and the price of labor rises. The ratio of the marginal product of labor to its price goes down, and the firm substitutes capital for labor. Similarly, an increase in the price of capital, all other things unchanged, would cause firms to substitute other factors of production for capital. The demand for capital, therefore, would fall.

Changes in Tax Policy

Government can indirectly affect the price of capital through changes in tax policy. For example, suppose the government enacts an investment tax credit for businesses, that is, a deduction of a certain percentage of their spending on capital from their profits before paying taxes. Such a policy would effectively lower the price of capital, causing firms to substitute capital for other factors of production and increasing the demand for capital. The repeal of an investment tax credit would lead to a decrease in the demand for capital.

306. Reading: Loanable Funds

The Market for Loanable Funds

When a firm decides to expand its capital stock, it can finance its purchase of capital in several ways. It might already have the funds on hand. It can also raise funds by selling shares of stock, as we discussed in a previous module. When a firm sells stock, it is selling shares of ownership of the firm. It can borrow the funds for the capital from a bank. Another option is to issue and sell its own bonds. A bond is a promise to pay back a certain amount at a certain time. When a firm borrows from a bank or sells bonds, of course, it accepts a liability—it must make interest payments to the bank or the owners of its bonds as they come due.

Regardless of the method of financing chosen, a critical factor in the firm's decision on whether to acquire and hold capital and on how to finance the capital is the interest rate. The role of the interest rate is obvious when the firm issues its own bonds or borrows from a bank. But even when the firm uses its own funds to purchase the capital, it is forgoing the option of lending those funds directly to other firms by buying their bonds or indirectly by putting the funds in bank accounts, thereby allowing the banks to lend the funds. The interest rate gives the opportunity cost of using funds to acquire capital rather than putting the funds to the best alternative use available to the firm.

The interest rate is determined in a market in the same way that the price of potatoes is determined in a market: by the forces of demand and supply. The market in which borrowers (demanders of funds) and lenders (suppliers of funds) meet is the loanable funds market.

We will simplify our model of the role that the interest rate plays in the demand for capital by ignoring differences in actual interest rates that specific consumers and firms face in the economy. For example, the interest rate on credit cards is higher than the mortgage rate of interest, and large, established companies can borrow funds or issue bonds at lower interest rates than new, start-up companies can. Interest rates that firms face depend on a variety of factors, such as riskiness of the loan, the duration of the loan, and the costs of administering the loan. However, since we will focus on general tendencies that cause interest rates to rise or fall and since the various interest rates in the economy tend to move up and down together, the conclusions we reach about the market for loanable funds and how firms and consumers respond to interest rate changes will still be valid.

The Demand for Loanable Funds

In the previous section we learned that a firm's decision to acquire and keep capital depends on the net present value of the capital in question, which in turn depends on the interest rate. The lower the interest rate, the greater the amount of capital that firms will want to acquire and hold, since lower interest rates translate into more capital with positive net present values. The desire for more capital means, in turn, a desire for more loanable funds. Similarly, at higher interest rates, less capital will be demanded, because more of the capital in question will have negative net present values. Higher interest rates therefore mean less funding demanded.

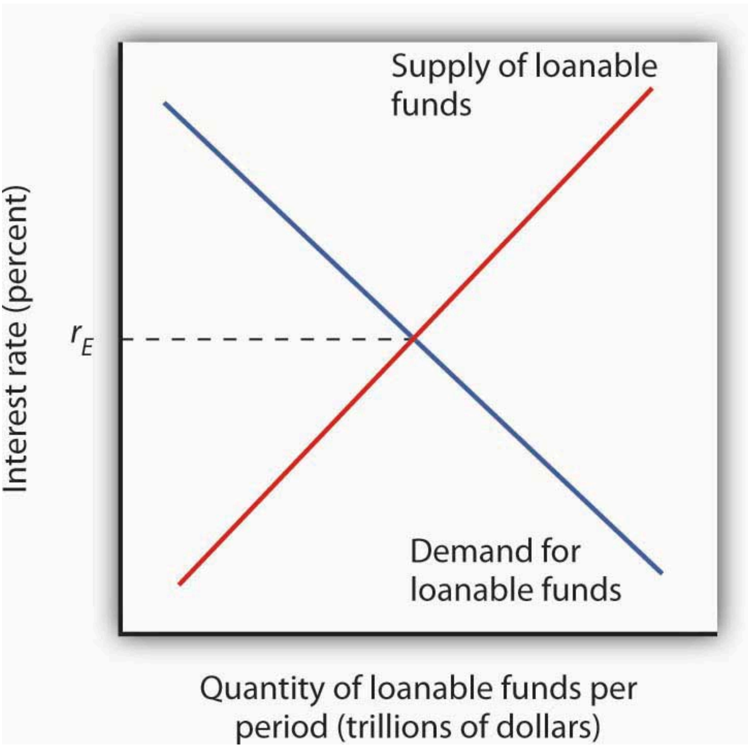


Figure 13.2. *The Demand and Supply of Loanable Funds. At lower interest rates, firms demand more capital and therefore more loanable funds. The demand for loanable funds is downward-sloping. The supply of loanable funds is generally upward-sloping. The equilibrium interest rate, rE, will be found where the two curves intersect.*

Thus the demand for loanable funds is downward-sloping, like the demand for virtually everything else, as shown in Figure 13.2. The lower the interest rate, the more capital firms will demand. The more capital that firms demand, the greater the funding that is required to finance it.

The Supply of Loanable Funds

Lenders are consumers or firms that decide that they are willing to forgo some current use of their funds in order to have more available in the future. Lenders supply funds to the loanable funds market. In general, higher interest rates make the lending option more attractive.

For consumers, however, the decision is a bit more complicated than it is for firms. In examining consumption choices across time, economists think of consumers as having an expected stream of income over their lifetimes. It is that expected income that defines their consumption possibilities. The problem for consumers is to determine when to consume this income. They can spend less of their projected income now and thus have more available in the future. Alternatively, they can boost their current spending by borrowing against their future income.

Saving is income not spent on consumption. (We shall ignore taxes in this analysis.) Dissaving occurs when consumption exceeds income during a period. Dissaving means that the individual's saving is negative. Dissaving can be financed either by borrowing or by using past savings. Many people, for example, save in preparation for retirement and then dissave during their retirement years.

Saving adds to a household's wealth. Dissaving reduces it. Indeed, a household's wealth is the sum of the value of all past saving less all past dissaving.

We can think of saving as a choice to postpone consumption. Because interest rates are a payment paid to people who postpone their use of wealth, interest rates are a kind of reward paid to savers. Will higher interest rates encourage the behavior they reward? The answer is a resounding "maybe." Just as higher wages might not increase the quantity of labor supplied, higher interest rates might not increase the quantity of saving. The problem, once again, lies in the fact that the income and substitution effects of a change in interest rates will pull in opposite directions.

Consider a hypothetical consumer, Tom Smith. Let us simplify the analysis of Mr. Smith's choices concerning the timing of consumption by assuming that there are only two periods: the present period is period 0, and the next is period 1. Suppose the interest rate is 8% and his income in both periods is expected to be $30,000.

Mr. Smith could, of course, spend $30,000 in period 0 and $30,000 in period 1. In that case, his saving equals zero in both periods. But he has alternatives. He could, for example, spend more than $30,000 in period 0 by borrowing against his income for period 1. Alternatively, he could spend less than $30,000 in period 0 and use his saving—and the interest he earns on that saving—to boost his consumption in period 1. If, for example, he spends $20,000 in period 0, his saving in period 0 equals $10,000. He will earn $800 interest on that saving, so he will have $40,800 to spend in the next period.

Suppose the interest rate rises to 10%. The increase in the interest rate has boosted the price of current consumption. Now for every $1 he spends in period 0 he gives up $1.10 in consumption in period 1, instead of $1.08, which was the amount that would have been given up in consumption in period 1 when the interest rate was 8%. A higher price produces a substitution effect that reduces an activity—Mr. Smith will spend less in the current period due to the substitution effect. The substitution effect of a higher interest rate thus boosts saving. But the higher interest rate also means that he earns more income on his saving. Consumption in the current period is a normal good, so an increase in income can be expected to increase current consumption. But an increase in current consumption implies a reduction in saving. The income effect of a higher interest rate thus tends to reduce saving. Whether Mr. Smith's savings will rise or fall in response to a higher interest rate depends on the relative strengths of the substitution and income effects.

To see how an increase in interest rates might reduce saving, imagine that Mr. Smith has decided that his goal is to have $40,800 to spend in period 1. At an interest rate of 10%, he can reduce his saving below $10,000 and still achieve his goal of having $40,800 to spend in the next period. The income effect of the increase in the interest rate has reduced his saving, and consequently his desire to supply funds to the loanable funds market.

Because changes in interest rates produce substitution and income effects that pull saving in opposite directions, we cannot be sure what will happen to saving if interest rates change. The combined effect of all consumers' and firms' decisions, however, generally leads to an upward-sloping supply curve for loanable funds, as shown in Figure 13.2. That is, the substitution effect usually dominates the income effect.

The equilibrium interest rate is determined by the intersection of the demand and supply curves in the market for loanable funds.

Capital and the Loanable Funds Market

If the quantity of capital demanded varies inversely with the interest rate, and if the interest rate is determined in the loanable funds market, then it follows that the demand for capital and the loanable funds market are interrelated. Because the acquisition of new capital is generally financed in the loanable funds market, a change in the demand for capital leads to a change in the demand for loanable funds—and that affects the interest rate. A change in the interest rate, in turn, affects the quantity of capital demanded on any demand curve.

The relationship between the demand for capital and the loanable funds market thus goes both ways. Changes in the demand for capital affect the loanable funds market, and changes in the loanable funds market can affect the quantity of capital demanded.

Changes in the Demand for Capital and the

Loanable Funds Market

Figure 13.3 suggests how an increased demand for capital by firms will affect the loanable funds market, and thus the quantity of capital firms will demand. In Panel (a) the initial interest rate is $r1$. At $r1$ in Panel (b) $K1$units of capital are demanded (on curve $D1$). Now suppose an improvement in technology increases the marginal product of capital, shifting the demand curve for capital in Panel (b) to the right to $D2$. Firms can be expected to finance the increased acquisition of capital by demanding more loanable funds, shifting the demand curve for loanable funds to $D2$ in Panel (a). The interest rate thus rises to $r2$. Consequently, in the market for capital the demand for capital is greater and the interest rate is higher. The new quantity of capital demanded is $K2$on demand curve $D2$.

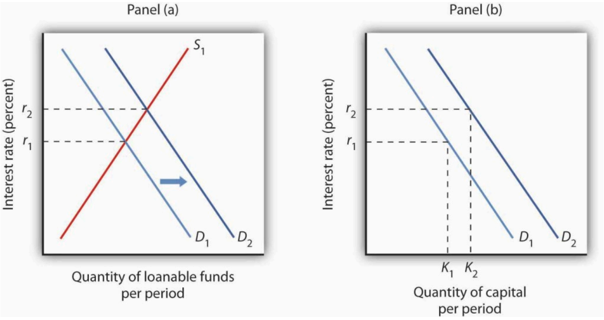


Figure 13.3. *Loanable Funds and the Demand for Capital. The interest rate is determined in the loanable funds market, and the quantity of capital demanded varies with the interest rate. Thus, events in the loanable funds market and the demand for capital are interrelated. If the demand for capital increases to D2 in Panel (b), the demand for loanable funds is likely to increase as well. Panel (a) shows the result in the loanable funds market—a shift in the demand curve for loanable funds from D1 to D2 and an increase in the*

interest rate from r1 to r2. At r2, the quantity of capital demanded will be K2, as shown in Panel (b).

Changes in the Loanable Funds Market and the Demand for Capital

Events in the loanable funds market can also affect the quantity of capital firms will hold. Suppose, for example, that consumers decide to increase current consumption and thus to supply fewer funds to the loanable funds market at any interest rate. This change in consumer preferences shifts the supply curve for loanable funds in Panel (a) of Figure 13.4 from S1 to S2 and raises the interest rate to *r*2. If there is no change in the demand for capital *D*1, the quantity of capital firms demand falls to *K*2 in Panel (b).

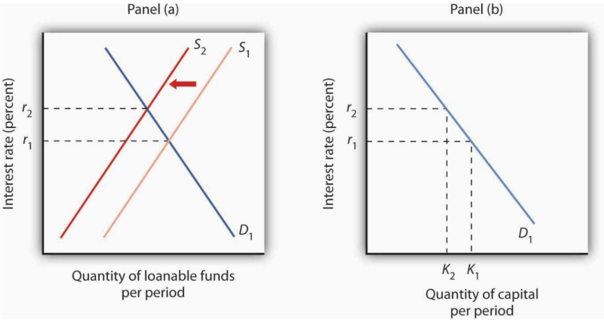


Figure 13.4. *A Change in the Loanable Funds Market and the Quantity of Capital Demanded. A change that begins in the loanable funds market can affect the quantity of capital firms demand. Here, a decrease in consumer saving causes a shift in the supply of loanable funds from S1 to S2 in Panel (a). Assuming there is no change in the demand for capital, the quantity of capital demanded falls from K1 to K2 in Panel (b).*

Our model of the relationship between the demand for capital and the loanable funds market thus assumes that the interest rate is determined in the market for loanable funds. Given the demand curve for capital, that interest rate then determines the quantity of capital firms demand.

Table 13.2 shows that a change in the quantity of capital that firms demand can begin with a change in the demand for capital or with a change in the demand for or supply of loanable funds. A change in the demand for capital affects the demand for loanable funds and hence the interest rate in the loanable funds market. The change in the interest rate leads to a change in the quantity of capital demanded. Alternatively, a change in the loanable funds market, which leads to a change in the interest rate, causes a change in quantity of capital demanded.

Table 13.2 Two Routes to Changes in the Quantity of Capital Demanded

A change originating in the capital market	A change originating in the loanable funds market
1. A change in the demand for capital leads to…	1. A change in the demand for or supply of loanable funds leads to …
2.…a change in the demand for loanable funds, which leads to…	2.…a change in the interest rate, which leads to…
3.…a change in the interest rate, which leads to…	3.…a change in the quantity of capital demanded.
4.…a change in the quantity of capital demanded.	

A change in the quantity of capital that firms demand can begin with a change in the demand for capital or with a change in the demand or supply of loanable funds.

KEY TAKEAWAYS

1. The net present value (NPV) of an investment project is equal to the present value of its expected revenues minus the present value of its expected costs. Firms will want to

undertake those investments for which the NPV is greater than or equal to zero.
2. The demand curve for capital shows that firms demand a greater quantity of capital at lower interest rates. Among the forces that can shift the demand curve for capital are changes in expectations, changes in technology, changes in the demands for goods and services, changes in relative factor prices, and changes in tax policy.
3. The interest rate is determined in the market for loanable funds. The demand curve for loanable funds has a negative slope; the supply curve has a positive slope.
4. Changes in the demand for capital affect the loanable funds market, and changes in the loanable funds market affect the quantity of capital demanded.

Case in Point: The Net Present Value of an MBA

An investment in human capital differs little from an investment in capital—one acquires an asset that will produce additional income over the life of the asset. One's education produces—or it can be expected to produce—additional income over one's working career.

Ronald Yeaple, a professor at the University of Rochester business school, has estimated the net present value (NPV) of an MBA obtained from each of 20 top business schools. The costs of attending each school included tuition and forgone income. To estimate the marginal revenue product of a degree, Mr. Yeaple started with survey data showing what graduates of each school were earning five years after obtaining their MBAs. He then estimated what students with the ability to attend those schools would have been earning without an MBA. The estimated marginal revenue product for each year is the difference between the salaries students earned with a degree versus what they would have earned without it. The NPV is then computed using

$$NPV_0 = R_0 - C_0 + \frac{R_1 - C_1}{1 + r} + \ldots + \frac{R_n - C_n}{(1 + r)^n}$$

The estimates given here show the NPV of an MBA over the first seven years of work after receiving the degree. They suggest that an MBA from 15 of the schools ranked is a good investment—but that a degree at the other schools might not be. Mr. Yeaple says that extending income projections beyond seven years would not significantly affect the analysis, because present values of projected income differentials with and without an MBA become very small.

While the Yeaple study is somewhat dated, a 2002 study by Stanford University Graduate School of Business professor Jeffrey Pfeffer and Stanford Ph.D. candidate Christina T. Fong reviewed 40 years of research on this topic and reached the conclusion that, "For the most part, there is scant evidence that the MBA credential, particularly from non-elite schools…are related to either salary or the attainment of higher level positions in organizations."

Of course, these studies only include financial aspects of the investment and did not cover any psychic benefits that MBA recipients may incur from more interesting work or prestige.

School	Net present value, first 7 years of work	School	Net present value, first 7 years of work
Harvard	$148,378	Virginia	$30,046
Chicago	106,847	Dartmouth	22,509
Stanford	97,462	Michigan	21,502
MIT	85,736	Carnegie-Mellon	18,679
Yale	83,775	Texas	17,459
Wharton	59,486	Rochester	−307
UCLA	55,088	Indiana	−3,315
Berkeley	54,101	NYU	−3,749
Northwestern	53,562	South Carolina	−4,565
Cornell	30,874	Duke	−17,631

307. Reading: Natural Resources and Conservation

Natural Resources and Conservation

Natural resources are the gifts of nature. They include everything from oil to fish in the sea to magnificent scenic vistas. The stock of a natural resource is the quantity of the resource with which the earth is endowed. For example, a certain amount of oil lies in the earth, a certain population of fish live in the sea, and a certain number of acres make up an area such as Yellowstone National Park or Manhattan. These stocks of natural resources, in turn, can be used to produce a flow of goods and services. Each year, we can extract a certain quantity of oil, harvest a certain quantity of fish, and enjoy a certain number of visits to Yellowstone.

As with capital, we examine the allocation of natural resources among alternative uses across time. By definition, natural resources cannot be produced. Our consumption of the services of natural resources in one period can affect their availability in future periods. We must thus consider the extent to which the expected demands of future generations should be taken into account when we allocate natural resources.

Natural resources often present problems of property rights in their allocation. A resource for which exclusive property rights have not been defined will be allocated as a common property resource. In such a case, we expect that the marketplace will not generate incentives to use the resource efficiently. In the absence of government intervention, natural resources that are common property may be destroyed. In this section, we shall consider natural resources for which exclusive property rights have been defined. The public sector's role in the allocation of common property resources is investigated in the module on the environment.

We can distinguish two categories of natural resources, those that are renewable and those that are not. A renewable natural resource is one whose services can be used in one period without necessarily reducing the stock of the resource that will be available in subsequent periods. The fact that they *can* be used in such a manner does not mean that they *will* be; renewable natural resources can be depleted. Wilderness areas, land, and water are renewable natural resources. The consumption of the services of an exhaustible natural resource, on the other hand, necessarily reduces the stock of the resource. Oil and coal are exhaustible natural resources.

Exhaustible Natural

Resources

Owners of exhaustible natural resources can be expected to take the interests of future as well as current consumers into account in their extraction decisions. The greater the expected future demand for an exhaustible natural resource, the greater will be the quantity preserved for future use.

Expectations and Resource Extraction

Suppose you are the exclusive owner of a deposit of oil in Wyoming. You know that any oil you pump from this deposit and sell cannot be replaced. You are aware that this is true of all the world's oil; the consumption of oil inevitably reduces the stock of this resource.

If the quantity of oil in the earth is declining and the demand for this oil is increasing, then it is likely that the price of oil will rise in the future. Suppose you expect the price of oil to increase at an annual rate of 15%.

Given your expectation, should you pump some of your oil out of the ground and sell it? To answer that question, you need to know the interest rate. If the interest rate is 10%, then your best alternative is to leave your oil in the ground. With oil prices expected to rise 15% per year, the dollar value of your oil will increase faster if you leave it in the ground than if you pump it out, sell it, and purchase an interest-earning asset. If the market interest rate were greater than 15%, however, it would make sense to pump the oil and sell it now and use the revenue to purchase an interest-bearing asset. The return from the interest-earning asset, say 16%, would exceed the 15% rate at which you expect the value of your oil to increase. Higher interest rates thus reduce the willingness of resource owners to preserve these resources for future use.

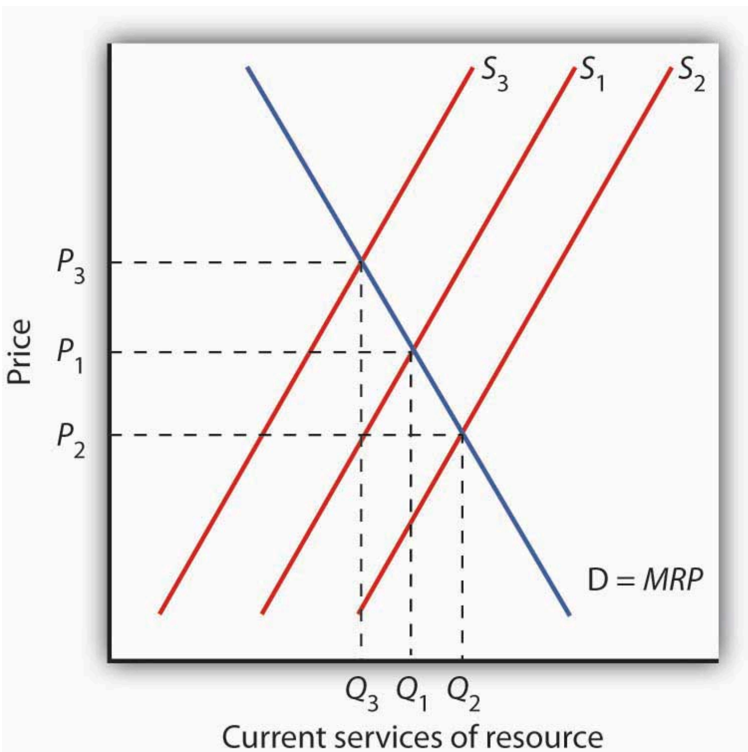


Figure 13.8. *Future Generations and Exhaustible Natural Resources. The current demand D for services of an exhaustible resource is given by the marginal revenue product (MRP). S1 reflects the current marginal cost of extracting the resource, the prevailing interest rate, and expectations of future demand for the resource. The level of current consumption is thus at Q1. If the interest rate rises, the supply curve shifts to S2, causing the price of the resource to fall to P2 and the quantity consumed to rise to Q2. A drop in the interest rate shifts the supply curve to S3, leading to an increase in price to P3 and a decrease in consumption to Q3.*

The supply of an exhaustible resource such as oil is thus governed by its current price, its expected future price, and the interest rate. An increase in the expected future price—or a reduction in the interest rate—reduces the supply of oil today, preserving more for future use. If owners of oil expect lower prices in the future, or if the interest rate rises, they will supply more oil today and conserve less for future use. This relationship is illustrated in Figure 13.8 "Future Generations and Exhaustible Natural Resources". The current demand *D* for these services is given by their marginal revenue product (*MRP*). Suppose S_1 reflects the current marginal cost of extracting the resource, the prevailing interest rate, and expectations of future demand for the resource. If the interest rate increases, owners will be willing to supply more of the natural resource at each price, thereby shifting the supply curve to the right to S_2. The current price of the resource will fall. If the interest rate falls, the supply curve for the resource will shift to the left to S_3 as more owners of the resource decide to leave more of the resource in the earth. As a result, the current price rises.

Resource Prices Over Time

Since using nonrenewable resources would seem to mean exhausting a fixed supply, then one would expect the prices of exhaustible natural resources to rise over time as the resources become more and more scarce. Over time, however, the prices of most exhaustible natural resources have fluctuated considerably relative to the prices of all other goods and services. Figure 13.9 "Natural Resource Prices, 1980–2007" shows the prices of four major exhaustible natural resources from 1980 to 2007. Prices have been adjusted for inflation to reflect the prices of these resources relative to other prices.

During the final two decades of the twentieth century, exhaustible natural resource prices were generally falling or stable. With the start of the current century, their prices have been rising. In short, why do prices of natural resources fluctuate as they do? Should the process of continuing to “exhaust” them just drive their prices up over time?

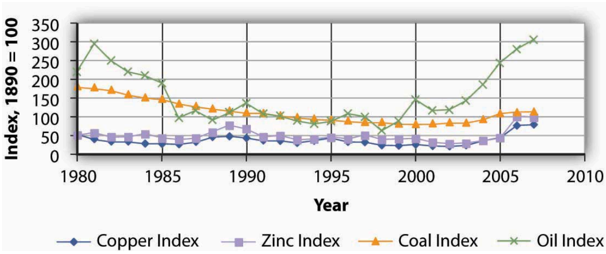


***Figure 13.9.** Natural Resource Prices, 1980–2007. The chart shows changes in the prices of five exhaustible resources—chromium, copper, nickel, tin, and tungsten (relative to the prices of other goods and services)—from 1890–2003. Sources: U.S. Bureau of the Census, Statistical Abstract of the United States, online; U.S. Energy Information Administration, Annual Energy Review, online.*

In setting their expectations, people in the marketplace must anticipate not only future demand but future supply as well. Demand in future periods could fall short of expectations if new technologies produce goods and services using less of a natural resource. That has clearly happened. The quantity of energy—which is generally produced using exhaustible fossil fuels—used to produce a unit of output has fallen by more than half in the last three decades. At the same time, rising income levels around the world, particularly in China and India over the last two decades, have led to increased demand for energy.

Supply increases when previously unknown deposits of natural resources are discovered and when technologies are developed to extract and refine resources more cheaply. Figure 13.10 "An Explanation for Falling Resource Prices" shows that discoveries that reduce the demand below expectations and increase the supply of natural resources can push prices down in a way that people in previous periods might not have anticipated. This scenario explains the fall in some prices of natural resources in the latter part of the twentieth century. To explain the recent rise in exhaustible natural resources prices, we can say that the factors contributing to increased demand for energy and some other exhaustible natural resources were outweighing the factors contributing to increased supply, resulting in higher prices—a scenario opposite to what is shown in Figure 13.10 "An Explanation for Falling Resource Prices". This upward trend began to reverse itself again in late 2008, as the world economies began to slump.

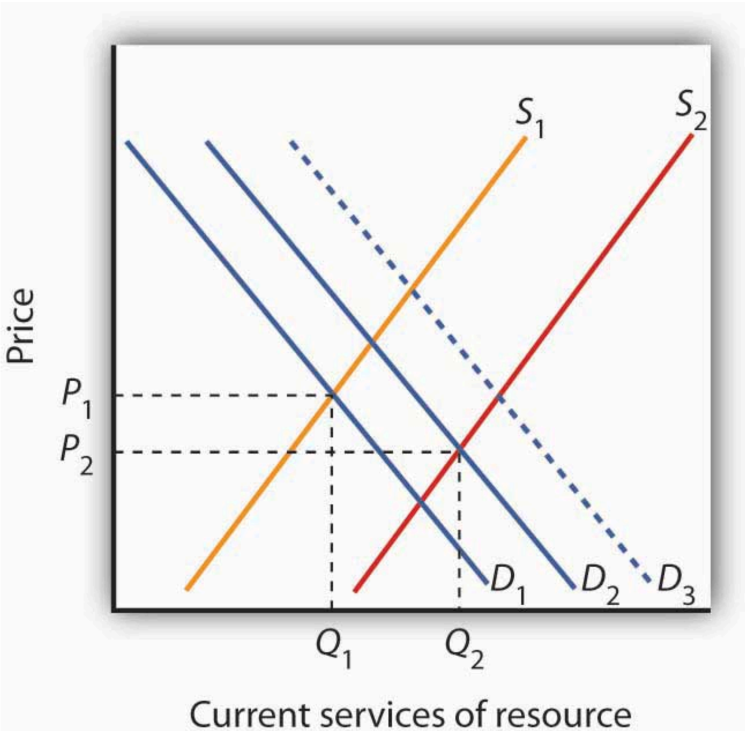


Figure 13.10. *An Explanation for Falling Resource Prices. Demand for resources has increased over time from D1 to D2, but this shift in demand is less than it would have been (D3) if technologies for producing goods and services using less resource per unit of output had not been developed. Supply of resources has increased from S1 to S2 as a result of the discovery of deposits of natural resources and/or development of new technologies for extracting and refining resources. As a result, the prices of many natural resources have fallen.*

Will we ever run out of exhaustible natural resources? Past experience suggests that we will not. If no new technologies or discoveries that reduce demand or increase supply occur, then resource prices will rise. As they rise, consumers of these resources will demand lower quantities of these resources. Eventually, the price of a particular resource could rise so high that the quantity demanded would fall to zero. At that point, no more of the resource would be used. There would still be some of the resource in the earth—it simply would not be practical to use more of it. The market simply will not allow us to "run out" of exhaustible natural resources.

Renewable Natural Resources

As is the case with exhaustible natural resources, our consumption of the services of renewable natural resources can affect future generations. Unlike exhaustible resources, however, renewable resources can be consumed in a way that does not diminish their stocks.

Carrying Capacity and Future Generations

The quantity of a renewable natural resource that can be consumed in any period without reducing the stock of the resource available in the next period is its carrying capacity. Suppose, for example, that a school of 10 million fish increases by 1 million fish each year. The carrying capacity of the school is therefore 1 million fish per year—the harvest of 1 million fish each year will leave the size of the population unchanged. Harvests that exceed a resource's carrying capacity reduce the stock of the resource; harvests that fall short of it increase that stock.

As is the case with exhaustible natural resources, future generations have a stake in current consumption of a renewable resource. Figure 13.11 "Future Generations and Renewable Resources" shows the efficient level of consumption of such a resource. Suppose Q_{cap} is the carrying capacity of a particular resource and S_1 is the supply curve that reflects the current marginal cost of utilizing the resource, including costs for the labor and capital required to make its services available, given the interest rate and expected future demand. The efficient level of consumption in the current period is found at point E, at the intersection of the current period's demand and supply curves. Notice that in the case shown, current consumption at Q_1 is less than the carrying capacity of the resource. A larger stock of this resource will be available in subsequent periods than is available now.

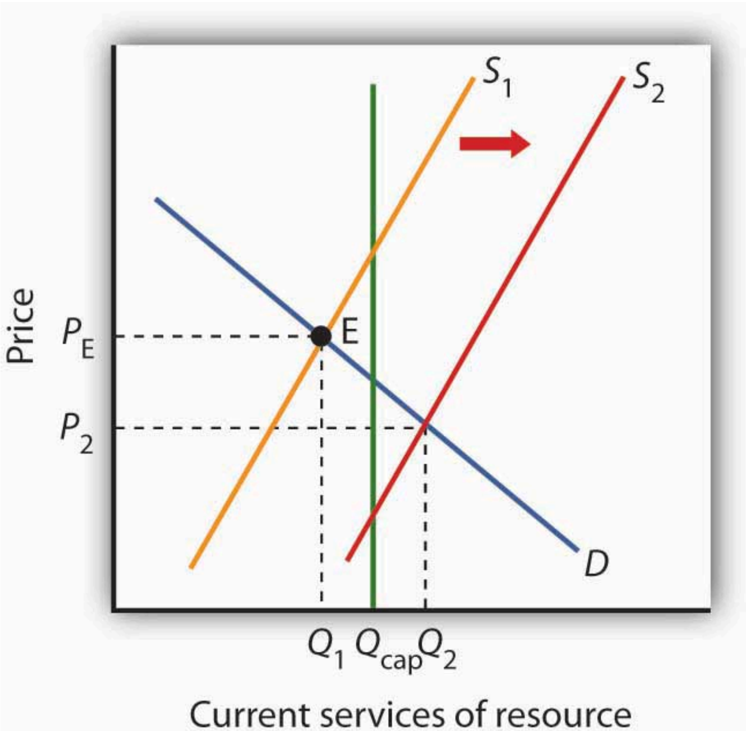


Figure 13.11. *Future Generations and Renewable Resources. The efficient quantity of services to consume is determined by the intersection S1 and the demand curve D. This intersection occurs at point E at a quantity of Q1. This lies below the carrying capacity Qcap. An increase in interest rates, however, shifts the supply curve to S2. The efficient level of current consumption rises to Q2, which now exceeds the carrying capacity of the resource.*

Now suppose interest rates increase. As with nonrenewable resources, higher interest rates shift the supply curve to the right, as shown by S_2. The result is an increase in current consumption to Q_2. Now consumption exceeds the carrying capacity, and the stock of the resource available to future generations will be reduced. While this solution may be efficient, the resource will not be sustained over time at current levels.

If society is concerned about a reduction in the amount of the resource available in the future, further steps may be required to preserve it. For example, if trees are being cut down faster than they are being replenished in a particular location, such as the Amazon in Brazil, a desire to maintain biological diversity might lead to conservation efforts.

Economic Rent and The Market for Land

We turn finally to the case of land that is used solely for the space it affords for other activities—parks, buildings, golf courses, and so forth. We shall assume that the carrying capacity of such land equals its quantity.

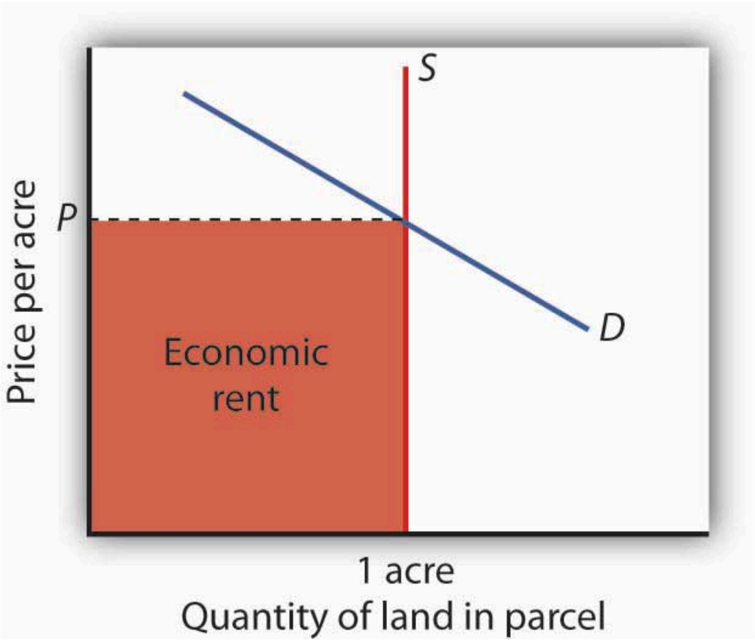


Figure 13.12. *The Market for Land. The price of a one-acre parcel of land is determined by the intersection of a vertical supply curve and the demand curve for the parcel. The sum paid for the parcel, shown by the shaded area, is economic rent.*

The supply of land is a vertical line. The quantity of land in a particular location is fixed. Suppose, for example, that the price of a one-acre parcel of land is zero. At a price of zero, there is still one acre of land; quantity is unaffected by price. If the price were to rise, there would still be only one acre in the parcel. That means that the price of the parcel exceeds the minimum price—zero—at which the land would be available. The amount by which any price exceeds the minimum price necessary to make a resource available is called economic rent.

The concept of economic rent can be applied to any factor of production that is in fixed supply above a certain price. In this sense, much of the salary received by Brad Pitt constitutes economic rent.

At a low enough salary, he might choose to leave the entertainment industry. How low would depend on what he could earn in a best alternative occupation. If he earns $30 million per year now but could earn $100,000 in a best alternative occupation, then $29.9 million of his salary is economic rent. Most of his current earnings are in the form of economic rent, because his salary substantially exceeds the minimum price necessary to keep him supplying his resources to current purposes.

KEY TAKEAWAYS

- Natural resources are either exhaustible or renewable.
- The demand for the services of a natural resource in any period is given by the marginal revenue product of those services.
- Owners of natural resources have an incentive to take into account the current price, the expected future demand for them, and the interest rate when making choices about resource supply.
- The services of a renewable natural resource may be consumed at levels that are below or greater than the carrying capacity of the resource.
- The payment for a resource above the minimum price necessary to make the resource available is economic rent.

Case in Point: World Oil Dilemma

The world is going to need a great deal more oil. Perhaps soon.

The International Energy Agency, regarded as one of the world's most reliable in assessing the global energy market, says that world oil production must increase from 87 million barrels per day in 2008 to 99 million barrels per day by 2015. Looking farther ahead, the situation gets scarier. Jad Mouawad reported in *The New York Times* that the number of cars and trucks in the world is expected to double—to 2 billion—in 30 years. The number of passenger jetliners in the world will double in 20 years. The IEA says that the demand for oil will increase by 35% by 2030. Meeting that demand would, according to the *Times*, require pumping an additional 11 billion barrels of oil each year—an increase of 13%.

Certainly some in Saudi Arabia, which holds a quarter of the world's oil reserves, were sure it would be capable of meeting the world's demand for oil, at least in the short term. In the summer of 2005, Peter Maass of *The New York Times* reported that Saudi Arabia's oil minister, Ali al-Naimi, gave an upbeat report in Washington, D.C. to a group of world oil officials. With oil prices then around $55 a barrel, he said, "I want to assure you here today that Saudi Arabia's reserves are plentiful, and we stand ready to increase output as the market dictates." The minister may well have been speaking in earnest. But, according to the U. S. Energy Information Administration, Saudi Arabia's oil production was 9.6 million barrels per day in 2005. It *fell* to 8.7 million barrels per day in 2006 and to 8.7 million barrels per day in 2007. The agency reports that world output also fell in each of those years. World oil prices soared to $147 per barrel in June of 2008. What happened?

Much of the explanation for the reduction in Saudi Arabia's output in 2006 and 2007 can be found in one field. More than half of the country's oil production comes from the Ghawar field, the most productive oil field in the world. Ghawar was discovered in 1948 and has provided the bulk of Saudi Arabia's oil. It has given the kingdom and the world more than 5 million barrels of oil per day for well over 50 years. It is, however, beginning to lose pressure. To continue getting oil from it, the Saudis have begun injecting the field with seawater. That creates new pressure and allows continued, albeit somewhat reduced, production. Falling production at Ghawar has been at the heart of Saudi Arabia's declining output.

The Saudi's next big hope is an area known as the Khurais complex. An area about half the size of Connecticut, the Saudis are counting on Khurais to produce 1.2 million barrels per day beginning in 2009. If it does, it will be the world's fourth largest oil field, behind Ghawar and fields in Mexico and Kuwait. Khurais, however, is no Ghawar. Not only is its expected yield much smaller, but it is going to be far more difficult to exploit. Khurais has no pressure of its own. To extract any oil from it, the Saudis will have to pump a massive amount of seawater from the Persian Gulf, which is 120 miles from Khurais. Injecting the water involves an extraordinary complex of pipes, filters, and more than 100 injection wells for the seawater. The whole project will cost a total of $15 billion. The Saudis told *The Wall Street Journal* that the development of the Khurais complex is the biggest industrial project underway in the world. The Saudis have used seismic technology to take more than 2.8 million 3-dimensional pictures of the deposit, trying to gain as complete an understanding of what lies beneath the surface as possible. The massive injection of seawater is risky. Done incorrectly, the introduction of the seawater could make the oil unusable.

Khurais illustrates a fundamental problem that the world faces as it contemplates its energy future. The field requires massive investment for an extraordinarily uncertain outcome, one that will only increase Saudi capacity from about 11.3 million barrels per day to 12.5.

Sadad al-Husseini, who until 2004 was the second in command at Aramco and is now a private energy consultant, doubts that Saudi

Arabia will be able to achieve even that increase in output. He says that is true of the world in general, that the globe has already reached the maximum production it will ever achieve—the so-called "peak production" theory. What we face, he told *The Wall Street Journal* in 2008, is a grim future of depleting oil resources and rising prices.

Rising oil prices, of course, lead to greater conservation efforts, and the economic slump that took hold in the latter part of 2008 has led to a sharp reversal in oil prices. But, if "peak production" theory is valid, lower oil prices will not persist after world growth returns to normal. This idea is certainly one to consider as we watch the path of oil prices over the next few years.

Self Check: Factors of Production

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the ten Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/699

308. Outcome: Competition and Wages

What you'll learn to do: analyze how perfect/imperfect competition between buyers and sellers of factors can impact wages, interest, and rents

In this section, you learn about markets with only one buyer (instead of seller), known as a monopsony, and what happens to prices and wages when buyers and sellers negotiate through collective bargaining.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Monopsony
- Reading: Monopoly and Monopsony: A Comparison
- Reading: Monopsony and the Minimum Wage
- Reading: Price Setters on the Supply Side
- Simulation: Income Distribution
- Self Check: Competition and Wages

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

309. Reading: Monopsony

Start Up: Hockey Players Frozen Out

On October 30, 2004, Columbus Blue Jackets' center Todd Marchant would ordinarily have been getting ready to open the 2004–2005 National Hockey League (NHL) season before a packed house in a game against the Dallas Stars in Dallas. Instead, he was home and devoting his season to coaching his six-year-old daughter's hockey team.

Mr. Marchant was home because the Commissioner of the NHL, Gary Bettman, had ordered players locked out on September 15, when training camp was scheduled to begin and when the contract between the NHL and the Players Association expired. Mr. Bettman had warned for five years that he would take the drastic action of shutting down the hockey season unless owners and players could agree on a system to limit player salaries. In the NHL, player salaries amounted to 75% of team revenues. By contrast, player salaries represented 64% of team revenues in the National Football League and 59% of revenues in the American Basketball Association. Mr. Bettman contended that the league's 30 franchises had lost a combined $500 million in the previous two years.

Players and owners alike had a great deal of money at stake. The NHL was selling 90% of its seats available during the regular season and generating $2.1 billion per year in revenues. "No one likes losing money, but this year everyone involved in hockey may be losing something," Mr. Marchant told *Business Week*. Mr. Marchant lost $2.9 million as a result of the lockout.

Mr. Bettman and the owners were holding out for a "salary cap" that would limit player salaries to 53% of team revenues. According to Mark Hyman of *Business Week*, that would reduce average salaries in hockey from $1.8 million to $1.3 million. "We're not going to play under a salary cap; we're dead set against it," Brad Lucovich, defenseman for Dallas, told *Business Week*. But the owners were similarly adamant. They were perfectly willing to forego revenues from the season—and to avoid paying player salaries—to establish a salary cap.

Were the owners being greedy? Or were the players at fault? For economists, the notions of "greed" or "blame" were not the issue. Economists assume that all individuals act in their own self-interest. In the case of the hockey lockout, which eliminated the 2004–05 season, players and owners were in a face-off in which a great deal of money was at stake. Owners had tried to establish a cap in 1994; the resulting labor dispute shut down half the season. Ultimately, the players prevailed and no caps were imposed. The 2005 lockout ended in nearly the opposite way. In the new contract, player salaries are capped and may not exceed 54% of league revenues. To most observers, it seemed that the team owners had won this battle.

Revolutionary changes in the rules that govern relations between the owners of sports teams and the players they hire have produced textbook examples of the economic forces at work in the determination of wages in imperfectly competitive markets. Markets for labor and other factors of production can diverge from the conditions of perfect competition in several ways, all of which involve price-setting behavior. Firms that purchase inputs may be price setters. Suppliers of inputs may have market power as well: a firm may have monopoly control over some key input or input suppliers may band together to achieve market power. Workers may organize unions. Suppliers of services, such as physicians and hairdressers, have formed associations that exert power in the marketplace.

This section applies the marginal decision rule to the analysis of imperfectly competitive markets for labor and other factors of production. Imperfect competition in these markets generally results in a reduction in the quantity of an input used, relative to the competitive equilibrium. The price of the input, however, could be higher or lower than in perfect competition, depending on the nature of the market structure involved.

Price-Setting Buyers: The Case of Monopsony

We have seen that market power in product markets exists when firms have the ability to set the prices they charge, within the limits of the demand curve for their products. Depending on the factor supply curve, firms may also have some power to set prices they pay in factor markets.

A firm can set price in a factor market if, instead of a market-determined price, it faces an upward-sloping supply curve for the factor. This creates a fundamental difference between price-taking and price-setting firms in factor markets. A price-taking firm can hire any amount of the factor at the market price; it faces a horizontal supply curve for the factor at the market-determined price, as shown in Panel (a) of Figure 14.1. A price-setting firm faces an upward-sloping supply curve such as S in Panel (b). It obtains Q_1 units of the factor when it sets the price P_1. To obtain a larger quantity, such as Q_2, it must offer a higher price, P_2.

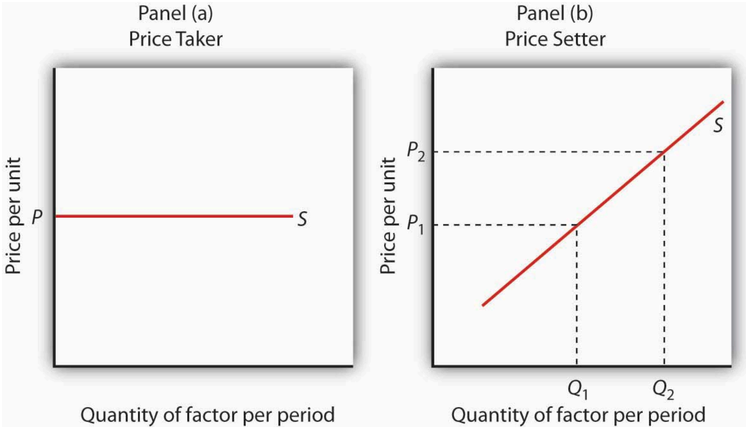


Figure 14.1. *Factor Market Price Takers and Price Setters. A price-taking firm faces the market-determined price P for the factor in Panel (a) and can purchase any quantity it wants at that price. A price-setting firm faces an upward-sloping supply curve S in Panel (b). The price-setting firm sets the price consistent with the quantity of the factor it wants to obtain. Here, the firm can obtain Q1 units at a price P1, but it must pay a higher price per unit, P2, to obtain Q2 units.*

Consider a situation in which one firm is the only buyer of a particular factor. An example might be an isolated mining town where the mine is the single employer. A market in which there is only one buyer of a good, service, or factor of production is called a monopsony. Monopsony is the buyer's counterpart of monopoly. Monopoly means a single seller; monopsony means a single buyer.

Assume that the suppliers of a factor in a monopsony market are price takers; there is perfect competition in factor supply. But a single firm constitutes the entire market for the factor. That means that the monopsony firm faces the upward-sloping market supply curve for the factor. Such a case is illustrated in Figure 14.2, where the price and quantity combinations on the supply curve for the factor are given in the table.

Price	$0	2	4	6	8	10
Quantity supplied	0	1	2	3	4	5
Total factor cost	$0	2	8	18	32	50
MFC		$2	6	10	14	18

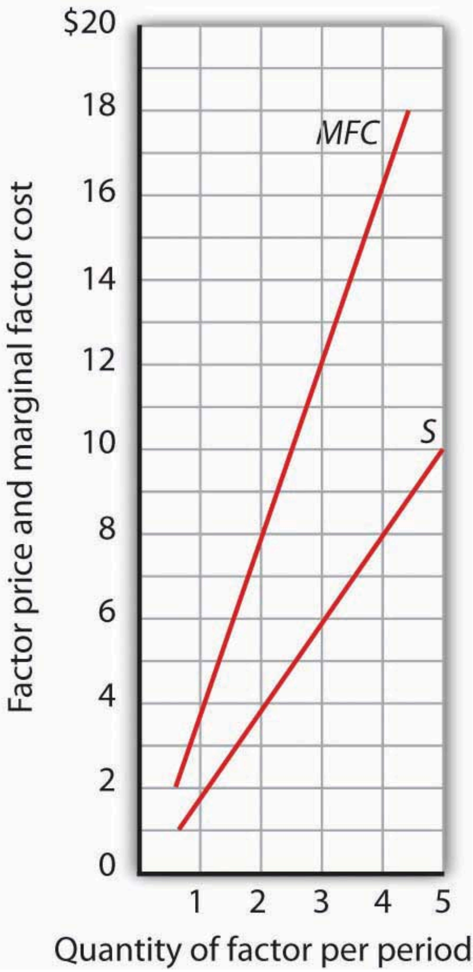


Figure 14.2. *Supply and Marginal Factor Cost. The table gives prices and quantities for the factor supply curve plotted in the graph. Notice that the marginal factor cost curve lies above the supply curve.*

Suppose the monopsony firm is now using three units of the factor at a price of $6 per unit. Its total factor cost is $18. Suppose the firm is considering adding one more unit of the factor. Given the supply curve, the only way the firm can obtain four units of the factor rather than three is to offer a higher price of $8 for all four units of the factor. That would increase the firm's total factor cost from $18 to $32. The marginal factor cost of the fourth unit of the factor is thus $14. It includes the $8 the firm pays for the fourth unit plus an additional $2 for each of the three units the firm was already using, since it has increased the prices for the factor to $8 from $6. The marginal factor cost (*MFC*) exceeds the price of the factor. We can plot the *MFC* for each increase in the quantity of the factor the firm uses; notice in Figure 14.2 that the *MFC* curve lies above the supply curve. As always in plotting in marginal values, we plot the $14 midway between units three and four because it is the increase in factor cost as the firm goes from three to four units.

Monopsony Equilibrium and the Marginal Decision Rule

The marginal decision rule, as it applies to a firm's use of factors, calls for the firm to add more units of a factor up to the point that the factor's *MRP* is equal to its *MFC*. Figure 14.3 illustrates this solution for a firm that is the only buyer of labor in a particular market.

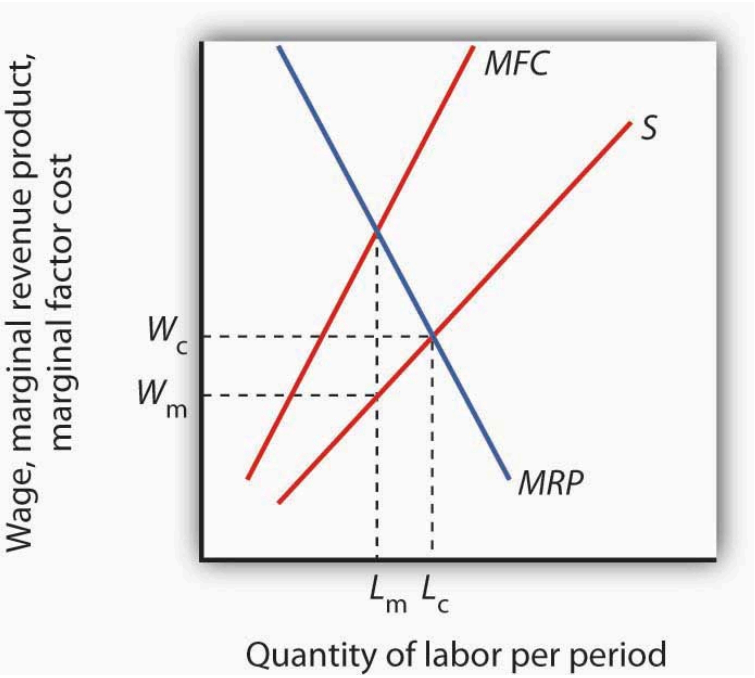


Figure 14.3. *Monopsony Equilibrium. Given the supply curve for labor, S, and the marginal factor cost curve, MFC, the monopsony firm will select the quantity of labor at which the MRP of labor equals its MFC. It thus uses Lm units of labor (determined by at the intersection of MRP and MFC) and pays a wage of Wm per unit (the wage is taken from the supply curve at which Lm units of labor are available). The quantity of labor used by the monopsony firm is less than would be used in a competitive market (Lc); the wage paid, Wm, is lower than would be paid in a competitive labor market.*

The firm faces the supply curve for labor, S, and the marginal factor cost curve for labor, *MFC*. The profit-maximizing quantity is determined by the intersection of the *MRP* and *MFC* curves—the firm will hire L_m units of labor. The wage at which the firm can obtain L_m units of labor is given by the supply curve for labor; it is W_m. Labor receives a wage that is less than its *MRP*.

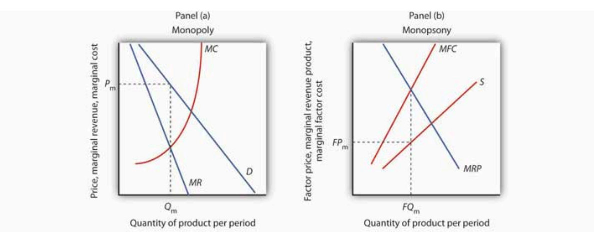
If the monopsony firm was broken up into a large number of small firms and all other conditions in the market remained unchanged, then the sum of the *MRP* curves for individual firms would be the market demand for labor. The equilibrium wage would be W_c, and the quantity of labor demanded would be L_c. Thus, compared to a competitive market, a monopsony solution generates a lower factor price and a smaller quantity of the factor demanded.

310. Reading: Monopoly and Monopsony: A Comparison

Monopoly and Monopsony: A Comparison

There is a close relationship between the models of monopoly and monopsony. A clear understanding of this relationship will help to clarify both models.

Figure 14.4 compares the monopoly and monopsony equilibrium solutions. Both types of firms are price setters: The monopoly is a price setter in its product market; the monopsony is a price setter in its factor market. Both firms must change price to change quantity: The monopoly must lower its product price to sell an additional unit of output, and the monopsony must pay more to hire an additional unit of the factor. Because both types of firms must adjust prices to change quantities, the marginal consequences of their choices are not given by the prices they charge (for products) or pay (for factors). For a monopoly, marginal revenue is less than price; for a monopsony, marginal factor cost is greater than price.



Market Choice	Monopoly	Monopsony
Nature of pricing choice	Price setter in product market.	Price setter in factor market.
Marginal consequences of choices	Selling an additional unit generates marginal revenue *MR*, which is less than *product price*.	Using an additional unit generates marginal *factor* cost *MFC*, which is greater than *factor price*.
Profit-maximizing solution	Produce where *MR* equals marginal cost *MC*—Q_m in Panel (a).	Use the factor at the quantity where *MRP* equals *MFC*—FQ_m in Panel (b).
Price	Determined by the point on the demand curve *D* that corresponds to the profit-maximizing quantity of product—P_m in Panel (a).	Determined by the point on the supply curve that corresponds to the profit-maximizing quantity of the factor—FP_m in Panel (b).

Figure 14.4 Monopoly and Monopsony. The graphs and the table provide a comparison of monopoly and monopsony.

Both types of firms follow the marginal decision rule: A monopoly produces a quantity of the product at which marginal revenue equals marginal cost; a monopsony employs a quantity of the factor at which marginal revenue product equals marginal factor cost. Both firms set prices at which they can sell or purchase the profit-maximizing quantity. The monopoly sets its product price based on the demand curve it faces; the monopsony sets its factor price based on the factor supply curve it faces.

Monopsony in the Real World

Although cases of pure monopsony are rare, there are many situations in which buyers have a degree of monopsony power. A buyer has monopsony power if it faces an upward-sloping supply curve for a good, service, or factor of production.

For example, a firm that accounts for a large share of employment in a small community may be large enough relative to the labor market that it is not a price taker. Instead, it must raise wages to attract more workers. It thus faces an upward-sloping supply curve and has monopsony power. Because buyers are more likely to have monopsony power in factor markets than in product markets, we shall focus on those.

The next section examines monopsony power in professional sports.

Monopsonies in Sports

Professional sports provide a setting in which economists can test theories of wage determination in competitive versus monopsony labor markets. In their analyses, economists assume professional teams are profit-maximizing firms that hire labor (athletes and other workers) to produce a product: entertainment bought by the fans who watch their games and by other firms that sponsor the games. Fans influence revenues directly by purchasing tickets and indirectly by generating the ratings that determine television and radio advertising revenues from broadcasts of games.

In a competitive system, a player should receive a wage equal to his or her *MRP*—the increase in team revenues the player is able to produce. As New York Yankees owner George Steinbrenner once put it, "You measure the value of a ballplayer by how many fannies he puts in the seats."

The monopsony model, however, predicts that players facing monopsony employers will receive wages that are less than their *MRP*s. A test of monopsony theory, then, would be to determine whether players in competitive markets receive wages equal to their *MRP*s and whether players in monopsony markets receive less.

Since the late 1970s, there has been a major shift in the rules that govern relations between professional athletes and owners of sports teams. The shift has turned the once monopsonistic market for professional athletes into a competitive one. Before 1977, for example, professional baseball players in the United States played under the terms of the "reserve clause," which specified that a player was "owned" by his team. Once a team had acquired a player's contract, the team could sell, trade, retain, or dismiss the player. Unless the team dismissed him, the player was unable to offer his services for competitive bidding by other teams. Moreover, players entered major league baseball through a draft that was structured so that only one team had the right to bid for any one player. Throughout a player's career, then, there was always only one team that could bid on him—each player faced a monopsony purchaser for his services to major league baseball.

Conditions were similar in other professional sports. Many studies have shown that the salaries of professional athletes in various team sports fell far short of their *MRP*s while monopsony prevailed.

When the reserve clauses were abandoned, players' salaries shot up—just as economic theory predicts. Because players could offer their services to other teams, owners began to bid for their services. Profit-maximizing owners were willing to pay athletes their *MRP*s. Average annual salaries for baseball players rose from about $50,000 in 1975 to nearly $1.4 million in 1997. Average annual player salaries in men's basketball rose from $109,000 in 1976 to $2.24 million in 1998. Football players worked under an almost pure form of monopsony until 1989, when a few players were allowed free agency status each year. In 1993, when 484 players were released to the market as free agents, those players received pay increases averaging more than 100%. Under the NFL collective bargaining agreement in effect in 1998, players could become unrestricted free agents if they had been playing for four years. There were 305 unrestricted free agents (out of a total player pool of approximately 1,700) that year. About half signed new contracts with their old teams while the other half signed with new teams. Table 14.1 illustrates the impact of free agency in four professional sports.

Table 14.1 The Impact of Free Agency

Player Salaries As Percentage of Team Revenues

	MLB	**NBA**	**NFL**	**NHL**
1970–73	15.9	46.1	34.4	21.3
1998	48.4	54.2	55.4	58.4

Free agency has increased player share of total revenues in each of the major men's team sports. Table 14.1 gives player salaries as a percentage of team revenues for major league baseball (MLB), the National Basketball Association (NBA), the National Football League (NFL) and the National Hockey League (NHL) during the 1970–1973 period that players in each league worked under monopsony conditions and in 1998, when players in each league had gained the right of free agency.

Given the dramatic impact on player salaries of more competitive markets for athletes, events such as the 2004–2005 lockout in hockey came as no surprise. The agreement between the owners of hockey teams and the players in 2005 to limit the total payroll of each team reinstates some of the old monopsony power of the owners. Players had a huge financial stake in resisting such attempts.

Monopsony in Other Labor Markets

A firm that has a dominant position in a local labor market may have monopsony power in that market. Even if a firm does not dominate the total labor market, it may have monopsony power over certain types of labor. For example, a hospital may be the only large employer of nurses in a local market, and it may have monopsony power in employing them.

Colleges and universities generally pay part-time instructors considerably less for teaching a particular course than they pay full-time instructors. In part, the difference reflects the fact that full-time faculty members are expected to have more training and are expected to contribute far more in other areas. But the monopsony model suggests an additional explanation.

Part-time instructors are likely to have other regular employment. A university hiring a local accountant to teach a section of accounting does not have to worry that that person will go to another state to find a better offer as a part-time instructor. For part-time teaching, then, the university may be the only employer in town—and thus able to exert monopsony power to drive the part-time instructor's wage below the instructor's *MRP*.

Monopsony in Other Factor Markets

Monopsony power may also exist in markets for factors other than labor. The military in different countries, for example, has considerable monopsony power in the market for sophisticated military goods. Major retailers often have some monopsony power with respect to some of their suppliers. Sears, for example, is the only wholesale buyer of Craftsman brand tools. One major development in medical care in recent years has been the emergence of managed care organizations that contract with a large number of employers to purchase medical services on behalf of employees. These organizations often have sufficient monopsony power to force down the prices charged by providers such as drug companies, physicians, and hospitals. Countries in which health care is provided by the government, such as Canada and the United Kingdom, are able to exert monopsony power in their purchase of health care services.

Whatever the source of monopsony power, the expected result is the same. Buyers with monopsony power are likely to pay a lower price and to buy a smaller quantity of a particular factor than buyers who operate in a more competitive environment.

KEY TAKEAWAYS

1. In the monopsony model there is one buyer for a good, service, or factor of production. A monopsony firm is a price setter in the market in which it has monopsony power.
2. The monopsony buyer selects a profit-maximizing solution by employing the quantity of factor at which marginal factor cost (*MFC*) equals marginal revenue product (*MRP*) and paying the price on the factor's supply curve corresponding to that quantity.

3. A degree of monopsony power exists whenever a firm faces an upward-sloping supply curve for a factor.

Case in Point: Professional Player Salaries and Monopsony

Professional athletes have not always enjoyed the freedom they have today to seek better offers from other teams. Before 1977, for example, baseball players could deal only with the team that owned their contract—one that “reserved” the player to that team. This reserve clause gave teams monopsony power over the players they employed. Similar restrictions hampered player mobility in men’s football, basketball, and hockey.

Gerald Scully, an economist at the University of Texas at Dallas, estimated the impact of the reserve clause on baseball player salaries. He sought to demonstrate that the player salaries fell short of *MRP*. Mr. Scully estimated the *MRP* of players in a two-step process. First, he studied the determinants of team attendance. He found that in addition to factors such as population and income in a team’s home city, the team’s win-loss record had a strong effect on attendance. Second, he examined the player characteristics that determined win-loss records. He found that for hitters, batting average was the variable most closely associated with a team’s winning percentage. For pitchers, it was the earned-run average—the number of earned runs allowed by a pitcher per nine innings pitched.

With equations that predicted a team’s attendance and its win-loss record, Mr. Scully was able to take a particular player, describe him by his statistics, and compute his *MRP*. Mr. Scully then subtracted costs associated with each player for such things as transportation, lodging, meals, and uniforms to obtain the player’s net *MRP*. He then compared players’ net *MRP*s to their salaries.

Mr. Scully’s results, displayed in the table below, show net *MRP* and salaries, estimated on a career basis, for players he classified as mediocre, average, and star-quality, based on their individual statistics. For average and star-quality players, salaries fell far below net *MRP*, just as the theory of monopsony suggests.

	Career Net *MRP*	**Career Salary**	**Salary As % of net *MRP***
Hitters			
Mediocre	−$129,300	$60,800	
Average	906,700	196,200	22
Star	3,139,100	477,200	15
Pitchers			
Mediocre	−53,600	54,800	
Average	1,119,200	222,500	20
Star	3,969,600	612,500	15

The fact that mediocre players with negative net *MRP*s received salaries presents something of a puzzle. One explanation could be that when they were signed to contracts, these players were expected to perform well, so their salaries reflected their expected contributions to team revenues. Their actual performance fell short, so their wages exceeded their *MRP*s. Another explanation could be that teams paid young players more than they were expected to contribute to revenues early in their careers in hopes that they would develop into profitable members of the team. In any event, Mr. Scully found that the costs of mediocre players exceeded their estimated contribution to team revenues, giving them negative net *MRP*s.

In 1977, a lawsuit filed by several baseball players resulted in the partial dismantling of the reserve clause. Players were given the right, after six years with a team, to declare themselves "free agents" and offer their services to other teams. Player salaries quickly rose. The accompanying table shows the pitchers that became free agents in 1977, their estimated net marginal revenue products, and their 1977 salaries. As you can see, salaries for pitchers came quite close to their net *MRPs*.

Pitcher	Net MRP	1977 Salary
Doyle Alexander	$166,203	$166,677
Bill Campbell	$205,639	$210,000
Rollie Fingers	$303,511	$332,000
Wayne Garland	$282,091	$230,000
Don Gullett	$340,846	$349,333

The same movement toward giving players greater freedom to deal with other teams occurred in the National Football League (NFL), the National Basketball Association (NBA), and the National Hockey League (NHL). The result in every case was the same: player salaries rose both in absolute terms and as a percentage of total team revenues. Table 14.1 gives player salaries as a percentage of total team revenues in the period 1970–73 and in 1998 for men's baseball (MLB), basketball, football, and hockey.

The greatest gains came in baseball, which had the most restrictive rules against player movement. Hockey players, too, ended up improving their salaries greatly. By 2004, their salaries totaled 75% of team revenues. The smallest gains came in basketball, where players already had options. The American Basketball Association was formed; it ultimately became part of the National Basketball Association. Basketball players also had the alternative of playing in Europe. But, the economic lesson remains clear: any weakening of the monopsony power of teams results in gains in player salaries.

311. Reading: Monopsony and the Minimum Wage

Monopsony and the Minimum Wage

We have seen that wages will be lower in monopsony than in otherwise similar competitive labor markets. In a competitive market, workers receive wages equal to their *MRP*s. Workers employed by monopsony firms receive wages that are less than their *MRP*s. This fact suggests sharply different conclusions for the analysis of minimum wages in competitive versus monopsony conditions.

In a competitive market, the imposition of a minimum wage above the equilibrium wage necessarily reduces employment, as we learned in the module on perfectly competitive labor markets. In a monopsony market, however, a minimum wage above the equilibrium wage could *increase* employment at the same time as it boosts wages!

Figure 14.5 shows a monopsony employer that faces a supply curve, *S*, from which we derive the marginal factor cost curve, *MFC*. The firm maximizes profit by employing *Lm* units of labor and paying a wage of $4 per hour. The wage is below the firm's *MRP*.

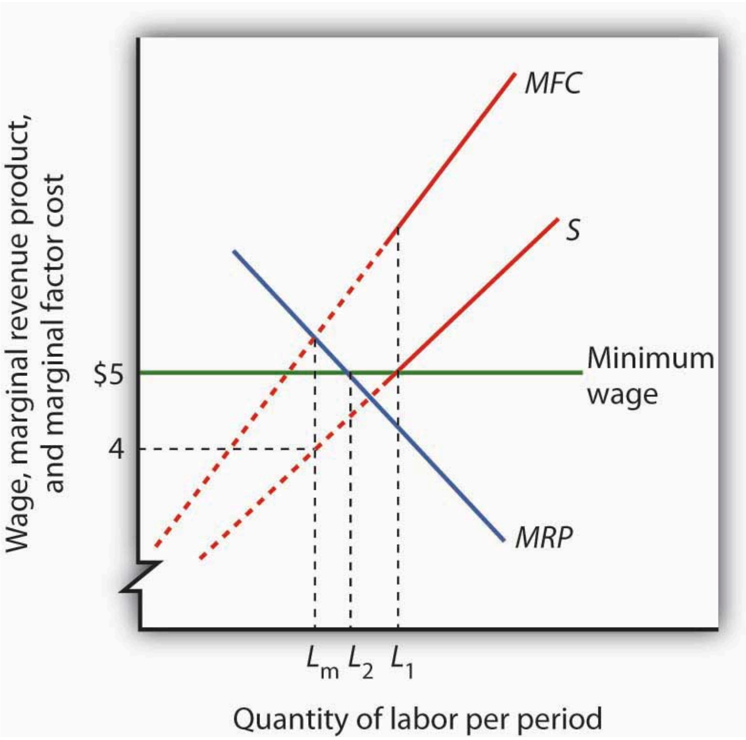


Figure 14.5. *Minimum Wage and Monopsony. A monopsony employer faces a supply curve S, a marginal factor cost curve MFC, and a marginal revenue product curve MRP. It maximizes profit by employing Lm units of labor and paying a wage of $4 per hour. The imposition of a minimum wage of $5 per hour makes the dashed sections of the supply and MFC curves irrelevant. The marginal factor cost curve is thus a horizontal line at $5 up to L1 units of labor. MRP and MFC now intersect at L2 so that employment increases.*

Now suppose the government imposes a minimum wage of $5 per hour; it is illegal for firms to pay less. At this minimum wage, L1units of labor are supplied. To obtain any smaller quantity of labor, the firm must pay the minimum wage. That means that the section of the supply curve showing quantities of labor supplied at wages below $5 is irrelevant; the firm cannot pay those wages. Notice that the section of the supply curve below $5 is shown as a dashed line. If the firm wants to hire more than *L*1units of labor, however, it must pay wages given by the supply curve.

Marginal factor cost is affected by the minimum wage. To hire additional units of labor up to *L*1, the firm pays the minimum wage. The additional cost of labor beyond *L*1continues to be given by the original *MFC* curve. The *MFC* curve thus has two segments: a horizontal segment at the minimum wage for quantities up to *L*1and the solid portion of the *MFC* curve for quantities beyond that.

The firm will still employ labor up to the point that *MFC* equals *MRP*. In the case shown in Figure 14.5, that occurs at *L*2. The firm thus increases its employment of labor in response to the minimum wage. This theoretical conclusion received apparent empirical validation in a study by David Card and Alan Krueger that suggested that an increase in New Jersey's minimum wage may have increased employment in the fast food industry. That conclusion became an important political tool for proponents of an increase in the minimum wage. The validity of those results has come under serious challenge, however, and the basic conclusion that a higher minimum wage would increase unemployment among unskilled workers in most cases remains the position of most economists. The discussion in the Case in Point summarizes the debate.

KEY TAKEAWAYS

1. In a competitive labor market, an increase in the minimum wage reduces employment and increases unemployment.
2. A minimum wage could increase employment in a monopsony labor market at the same time it increases wages.
3. Some economists argue that the monopsony model characterizes all labor markets and that this justifies a national increase in the minimum wage.
4. Most economists argue that a nationwide increase in the

minimum wage would reduce employment among low-wage workers.

Case in Point: The Monopsony-Minimum Wage Controversy

While the imposition of a minimum wage on a monopsony employer could increase employment and wages at the same time, the possibility is generally regarded as empirically unimportant, given the rarity of cases of monopsony power in labor markets. However, some studies have found that increases in the minimum wage have led to either increased employment or to no significant reductions in employment. These results appear to contradict the competitive model of demand and supply in the labor market, which predicts that an increase in the minimum wage will lead to a reduction in employment and an increase in unemployment.

The study that sparked the controversy was an analysis by David Card and Alan Krueger of employment in the fast food industry in Pennsylvania and New Jersey. New Jersey increased its minimum wage to $5.05 per hour in 1992, when the national minimum wage was $4.25 per hour. The two economists surveyed 410 fast food restaurants in the Burger King, KFC, Roy Rogers, and Wendy's chains just before New Jersey increased its minimum and again 10 months after the increase.

There was no statistically significant change in employment in the New Jersey franchises, but employment fell in the Pennsylvania franchises. Thus, employment in the New Jersey franchises "rose" relative to employment in the Pennsylvania franchises. Card and Krueger's results were widely interpreted as showing an increase in employment in New Jersey as a result of the increase in the minimum wage there.

Do minimum wages reduce employment or not? Some economists interpreted the Card and Krueger results as demonstrating widespread monopsony power in the labor market. Economist Alan Manning notes that the competitive model implies that a firm that pays a penny less than the market equilibrium wage will have zero employees. But, Mr. Manning notes that there are non-wage attributes to any job that, together with the cost of changing jobs, result in individual employers facing upward-sloping supply curves for labor and thus giving them monopsony power. And, as we have seen, a firm with monopsony power may respond to an increase in the minimum wage by increasing employment.

The difficulty with implementing this conclusion on a national basis is that, even if firms do have a degree of monopsony power, it is impossible to determine just *how much* power any one firm has and by how much the minimum wage could be increased for each firm. As a result, even if it were true that firms had such monopsony power, it would not follow that an increase in the minimum wage would be appropriate.

Even the finding that an increase in the minimum wage may not reduce employment has been called into question. First, there are many empirical studies that suggest that increases in the minimum wage do reduce employment. For example, a recent study of employment in the restaurant industry by Chicago Federal Reserve Bank economists Daniel Aaronson and Eric French concluded that a 10% increase in the minimum wage would reduce employment among unskilled restaurant workers by 2 to 4%. This finding was more in line with other empirical work. Further, economists point out that jobs have nonwage elements. Hours of work, working conditions, fellow employees, health insurance, and other fringe benefits of working can all be adjusted by firms in response to an increase in the minimum wage. Dwight Lee, an economist at the University of Georgia, argues that as a result, an increase in the minimum wage may not reduce employment but may reduce other fringe benefits that workers value more highly than wages themselves. So, an increase in the minimum wage may make even workers who receive higher wages worse off. One indicator that suggests that higher minimum wages may reduce the welfare of low income workers is that participation in the labor force by teenagers has been shown to fall as a result of higher minimum wages. If the opportunity to earn higher wages reduces the number of teenagers seeking those wages, it may indicate that low-wage work has become less desirable.

In short, the possibility that higher minimum wages might not reduce employment among low-wage workers does not necessarily mean that higher minimum wages improve the welfare of low income workers. Evidence that casts doubt on the proposition that higher minimum wages reduce employment does not remove many economists' doubt that higher minimum wages would be a good policy.

312. Reading: Price Setters on the Supply Side

Price Setters on the Supply Side

Buyers are not the only agents capable of exercising market power in factor-pricing choices. Suppliers of factor services can exercise market power and act as price setters themselves in two ways. First, a supplier may be a monopoly or have a degree of monopoly power in the supply of a factor. In that case, economists analyze the firm's choices as they would analyze those of any other imperfectly competitive firm. Second, individual suppliers of a factor of production may band together in an association to gain clout in the marketplace. Farmers, for example, often join forces to offset what they perceive as unfair market power on the part of buyers of their products. Workers may join together in a union in order to enhance their bargaining power with their employers. Each case is discussed below.

Monopoly Suppliers

A firm with monopoly power over a particular factor can be expected to behave like any other monopoly. It will choose its output where the marginal revenue and marginal cost curves intersect and charge a price taken from its demand curve. A monopoly supplier of a factor faces a demand curve that represents the *MRP* of the factor. This situation is illustrated in Figure 14.6. The firm will charge a price Pm equal to the *MRP* of the factor and sell Qm units of the factor.

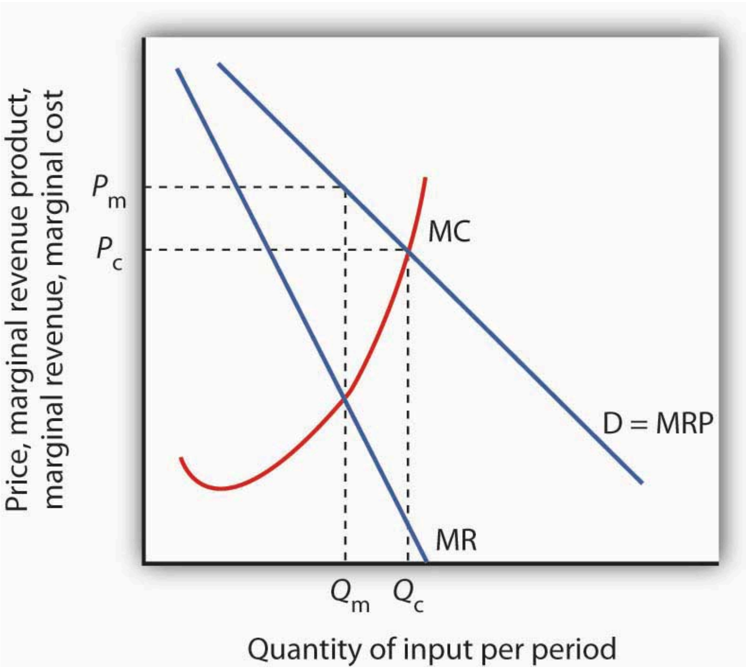


***Figure 14.6** Monopoly Factor Supply. A monopoly supplier of a factor of production acts just as any other monopoly firm. Here, the monopoly faces the demand curve D and the marginal revenue curve MR. Given the marginal cost curve MC, it maximizes profit by supplying Qm and charging a price Pm.*

Unions

A ***labor union*** is an organization of workers that negotiates with employers over wages and working conditions. A labor union seeks to change the balance of power between employers and workers by requiring employers to deal with workers collectively, rather than as individuals. Thus, negotiations between unions and firms are sometimes called ***collective bargaining***.

The subject of labor unions can be controversial. Supporters of labor unions view them as the workers' primary line of defense against efforts by profit-seeking firms to hold down wages. Critics of labor unions view them as having a tendency to grab as much as they can in the short term, even if it means injuring workers in the long run by driving firms into bankruptcy or by blocking the new technologies and production methods that lead to economic growth. We will start with some facts about union membership in the United States.

FACTS ABOUT UNION MEMBERSHIP AND PAY

According to the U.S. Bureau of Labor and Statistics, about 11.3% of all U.S. workers belong to unions. Following are some of the facts provided by the bureau for 2013:

- 12.0% of U.S. male workers belong to unions; 10.5% of female workers do
- 11.1% of white workers, 13.4 % of black workers, and 9.8 % of Hispanic workers belong to unions
- 12.5% of full-time workers and 6.0% of part-time workers are union members
- 4.2% of workers ages 16–24 belong to unions, as do 14% of workers ages 45-54
- Occupations in which relatively high percentages of workers belong to unions are the federal government (26.9% belong to a union), state government (31.3%), local government (41.7%); transportation and utilities (20.6%); natural resources, construction, and maintenance (16.3%); and production, transportation, and material moving (14.7%)

- Occupations that have relatively low percentages of unionized workers are agricultural workers (1.4%), financial services (1.1%), professional and business services (2.4%), leisure and hospitality (2.7%), and wholesale and retail trade (4.7%)

In summary, the percentage of workers belonging to a union is higher for men than women; higher for blacks than for whites or Hispanics; higher for the 45–64 age range; and higher among workers in government and manufacturing than workers in agriculture or service-oriented jobs. Table 15.2 lists the largest U.S. labor unions and their membership.

Table 15.2 The Largest American Unions in 2013(Source: U.S. Department of Labor, Bureau of Labor Statistics)

Union	Membership
National Education Association (NEA)	3.2 million
Service Employees International Union (SEIU)	2.1 million
American Federation of Teachers (AFT)	1.5 million
International Brotherhood of Teamsters (IBT)	1.4 million
The American Federation of State, County, and Municipal Workers (AFSCME)	1.3 million
United Food and Commercial Workers International Union	1.3 million
United Steelworkers	1.2 million
International Union, United Automobile, Aerospace and Agricultural Implement Workers of America (UAW)	990,000
International Association of Machinists and Aerospace Workers	720,000
International Brotherhood of Electrical Workers (IBEW)	675,000

In terms of pay, benefits, and hiring, U.S. unions offer a good news/bad news story. The good news for unions and their members is that their members earn about 20% more than nonunion workers, even after adjusting for factors such as years of work experience and education level. The bad news for unions is that the share of U.S. workers who belong to a labor union has been steadily declining for 50 years, as shown in Figure 15.2. About one-quarter of all U.S. workers belonged to a union in the mid-1950s, but only 11.3% of U.S. workers are union members today. If you leave out workers employed by the government (which includes teachers in public schools), only 6.6% of the workers employed by private firms now work for a union.

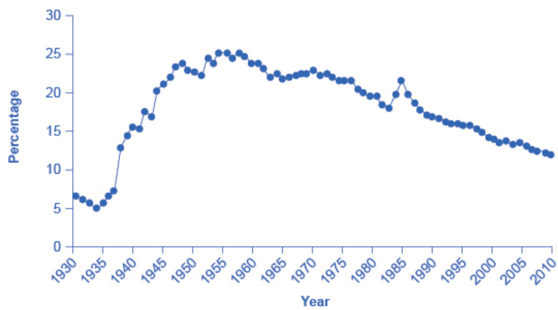


Figure 15.2. *Percentage of Wage and Salary Workers Who Are Union Members. The share of wage and salary workers who belong to unions rose sharply in the 1930s and 1940s, but has tailed off since then to 11.3% of all workers in 2012.*

A Brief History of Unions

Workers have united to try to better their lot at least since the Middle Ages, when the first professional guilds were formed in Europe. In the United States, "workingmen's societies" sprang up in the late eighteenth century. These organizations were craft unions uniting skilled workers in the same trade in an attempt to increase wages, shorten working hours, and regulate working conditions for their members.

One goal unions consistently sought was a closed shop, where only union members can be hired—an arrangement that gives unions monopoly power in the supply of labor. A second objective was to gain greater political and economic strength by joining together associations of different crafts. Union goals went largely unfulfilled until the twentieth century, when the courts began to favor collective bargaining between workers and employers in disputes over wages and working conditions. Closed-shop arrangements are illegal in the United States today, but many states permit union shop arrangements, in which a firm is allowed to hire nonunion workers who are required to join the union within a specified period. About 20 states have right-to-work laws which prohibit union shop rules.

The development of the industrial union, a form of union that represents the employees of a particular industry, regardless of their craft, also aided the growth of the labor movement. The largest industrial union in the United States, the AFL-CIO, was formed in 1955, when unions accounted for just over 35% of the labor force. The AFL-CIO remains an important economic and political force, but union strength has fallen since its peak in the 1950s; today, less than 10% of workers in the private sector belong to unions. Quite dramatically, in 2005, three unions, representing about a third of the total membership, withdrew from the AFL-CIO. The break-away unions argued that they would be more successful working on their own to recruit new members. The impact of this break-up will not be known for several years.

Part of the reason for the failure of unions to represent a larger share of workers lies in the market forces that govern wages. As the marginal revenue product of workers has risen throughout the economy, their wages have increased as well—whether they belonged to a union or not. Impressive economy-wide wage gains over the last two centuries may be one reason why the attraction of unions has remained weak.

HIGHER WAGES FOR UNION WORKERS

Why might union workers receive higher pay? What are the limits on how much higher pay they can receive? To analyze these questions, let's consider a situation where all firms in an industry must negotiate with a single union, and no *firm* is allowed to hire nonunion labor. If no labor union existed in this market, then equilibrium (E) in the labor market would occur at the intersection of the demand for labor (D) and the supply of labor (S) in Figure 15.3. The union can, however, threaten that, unless firms agree to the wages they demand, the workers will strike. As a result, the labor union manages to achieve, through negotiations with the firms, a union wage of Wu for its members, above what the equilibrium wage would otherwise have been.

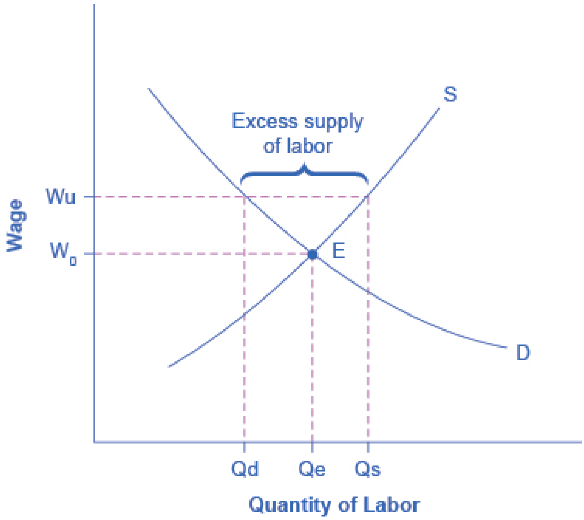


Figure 15.3. *Union Wage Negotiations. Without a union, the equilibrium at E would have involved the wage We and the quantity of labor Qe. However, the union is able to use its bargaining power to raise the wage to Wu. The result is an excess supply of labor for union jobs. That is, a quantity of labor supplied, Qs is greater than firms' quantity demanded for labor, Qd.*

This labor market situation resembles what a ***monopoly firm*** does in selling a product, but in this case a union is a monopoly selling labor to firms. At the higher union wage Wu, the firms in this industry will hire less labor than they would have hired in equilibrium. Moreover, an excess supply of workers want union jobs, but firms will not be hiring for such jobs.

From the union point of view, workers who receive higher wages are better off. However, notice that the quantity of workers (Qd) hired at the union wage Wu is smaller than the quantity Qe that would have been hired at the original equilibrium wage. A sensible union must recognize that when it pushes up the wage, it also reduces the incentive of firms to hire. This situation does not necessarily mean that union workers are fired. Instead, it may be that when union workers move on to other jobs or retire, they are not always replaced. Or perhaps when a firm expands production, it expands employment somewhat less with a higher union wage than it would have done with the lower equilibrium wage. Or perhaps a firm decides to purchase inputs from nonunion producers, rather than producing them with its own highly paid unionized workers. Or perhaps the firm moves or opens a new facility in a state or country where unions are less powerful.

From the firm's point of view, the key question is whether the higher wage of union workers is matched by higher productivity. If so, then the firm can afford to pay the higher union wages and, indeed, the demand curve for "unionized" labor could actually shift to the right. This could reduce the job losses as the equilibrium employment level shifts to the right and the difference between the equilibrium and the union wages will have been reduced. If worker unionization does not increase productivity, then the higher union wage will cause lower profits or losses for the firm.

Union workers might have higher productivity than nonunion workers for a number of reasons. First, higher wages may elicit higher productivity. Second, union workers tend to stay longer at a given job, a trend that reduces the employer's costs for training and hiring and results in workers with more years of experience. Many unions also offer job training and apprenticeship programs.

In addition, firms that are confronted with union demands for higher wages may choose production methods that involve more physical capital and less labor, resulting in increased labor productivity. Table 15.3 provides an example. Assume that a firm can produce a home exercise cycle with three different combinations of labor and manufacturing equipment. Say that labor is paid $16 an hour (including benefits) and the machines for manufacturing cost

$200 each. Under these circumstances, the total cost of producing a home exercise cycle will be lowest if the firm adopts the plan of 50 hours of labor and one machine, as the table shows. Now, suppose that a union negotiates a wage of $20 an hour including benefits. In this case, it makes no difference to the firm whether it uses more hours of labor and fewer machines or less labor and more machines, though it might prefer to use more machines and to hire fewer union workers. (After all, machines never threaten to strike—but they do not buy the final product or service either.) In the final column of the table, the wage has risen to $24 an hour. In this case, the firm clearly has an incentive for using the plan that involves paying for fewer hours of labor and using three machines. If management responds to union demands for higher wages by investing more in machinery, then union workers can be more productive because they are working with more or better physical capital equipment than the typical nonunion worker. However, the firm will need to hire fewer workers.

Table 15.3. Three Production Choices to Manufacture a Home Exercise Cycle

Hours of Labor	Number of Machines	Cost of Labor + Cost of Machine $16/hour	Cost of Labor + Cost of Machine $20/hour	Cost of Labor + Cost of Machine $24/hr
30	3	$480 + $600 = $1,080	$600 + $600 = $1,200	$720 + $600 = $1,320
40	2	$640 + $400 = $1,040	$800 + $400 = $1,200	$960 + $400 = $1,360
50	1	$800 + $200 = $1,000	$1,000 + $200 = $1,200	$1,200 + $200 = $1,400

In some cases, unions have discouraged the use of labor-saving physical capital equipment—out of the reasonable fear that new machinery will reduce the number of union jobs. For example, in 2002, the union representing longshoremen who unload ships and the firms that operate shipping companies and port facilities staged a work stoppage that shut down the ports on the western coast of the United States. A key issue in the dispute was the desire of the shipping companies and port operators to use handheld scanners for record-keeping and computer-operated cabs for loading and unloading ships—changes which the union opposed. President George W. Bush invoked the Labor Management Relations Act of 1947—commonly known as the *Taft-Hartley Act*—and asked a court to impose an 80-day "cooling-off period" in order to allow time for negotiations to proceed without the threat of a work stoppage. Federal mediators were called in, and the two sides agreed to a deal in November 2002. The ultimate agreement allowed the new technologies, but also kept wages, health, and pension benefits high for workers. In the past, presidential use of the Taft-Hartley Act sometimes has made labor negotiations more bitter and argumentative but, in this case, it seems to have smoothed the road to an agreement.

In other instances, unions have proved quite willing to adopt new technologies. In one prominent example, during the 1950s and 1960s, the *United Mineworkers union*demanded that mining companies install labor-saving machinery in the mines. The mineworkers' union realized that over time, the new machines would reduce the number of jobs in the mines, but the union leaders also knew that the mine owners would have to pay higher wages if the workers became more productive, and mechanization was a necessary step toward greater productivity.

In fact, in some cases union workers may be more willing to accept new technology than nonunion workers, because the union workers believe that the union will negotiate to protect their jobs and wages, whereas nonunion workers may be more concerned that the new technology will replace their jobs. In addition, union workers, who typically have higher job market experience and training, are likely to suffer less and benefit more than non-union workers from the introduction of new technology. Overall, it is hard to make a definitive case that union workers as a group are always either more or less welcoming to new technology than are nonunion workers.

Higher Wages and Other Union Goals

Higher wages once dominated the list of union objectives, but more recent agreements have also focused on nonwage issues involving job security, health insurance, provision of child care, and job safety. Unions such as the United Auto Workers have negotiated contracts under which members who are laid off will continue to receive payments nearly equal to the wages they earned while on the job. They have also pushed hard for retirement pensions and for greater worker involvement in management decisions.

Union efforts to obtain higher wages have different effects on workers depending on the nature of the labor market. When unions confront an employer with monopsony power, their task is clear: they seek a wage closer to *MRP* than the employer is paying. If the labor market is a competitive one in which wages are determined by demand and supply, the union's task is more difficult. Increasing the wage requires either increasing the demand for labor or reducing the supply. If the union merely achieves a higher wage in the absence of an increase in demand or a reduction in supply, then the higher wage will create a surplus of labor, or unemployment.

Increasing Demand

The demand for labor in a competitive market is found by summing the *MRP* curves of individual firms. Increasing demand thus requires increasing the marginal product of labor or raising the price of the good produced by labor.

One way that unions can increase the marginal product of their members is by encouraging investment in their human capital. Consequently, unions may pressure firms to implement training programs. Some unions conduct training efforts themselves.

Another way to increase the *MRP* of a factor is to reduce the use by firms of substitute factors. Unions generally represent skilled workers, and they are vigorous proponents of minimum wage laws that make unskilled workers more expensive. A higher minimum wage induces firms to substitute skilled for unskilled labor and thus increases the demand for the skilled workers unions represent.

Still another way to increase the *MRP* of labor is to increase the demand for the products labor produces. The form this union activity generally takes is in the promotion of "Made in the U.S.A." goods. Unions have also promoted restrictive trade legislation aimed at reducing the supply of foreign goods and thus increasing the demand for domestic ones.

Reducing Labor Supply

Unions can restrict the supply of labor in two ways. First, they can seek to slow the growth of the labor force; unions from the earliest times have aggressively opposed immigration. Union support for Social Security also cut the labor supply by encouraging workers to retire early. Second, unions can promote policies that make it difficult for workers to enter a particular craft. Unions representing plumbers and electrical workers, for example, have restricted the number of people who can enter these crafts in some areas by requiring that workers belong to a union and then limiting the union's membership.

Bilateral Monopoly

Suppose a union has negotiated a closed-shop arrangement (in a country where such arrangements are legal) with an employer that possesses monopsony power in its labor market. The union has a kind of monopoly in the supply of labor. A situation in which a monopsony buyer faces a monopoly seller is called bilateral monopoly. Wages in this model are indeterminate, with the actual wage falling somewhere between the pure monopoly and pure monopsony outcomes.

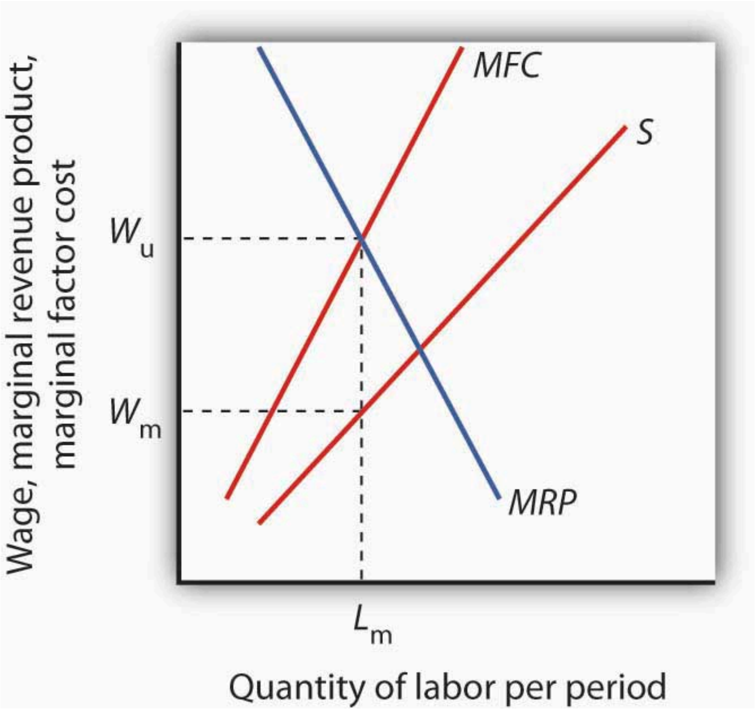


Figure 14.7 *Bilateral Monopoly. If the union has monopoly power over the supply of labor and faces a monopsony purchaser of the labor the union represents, the wage negotiated between the two will be indeterminate. The employer will hire Lm units of the labor per period. The employer wants a wage Wm on the supply curve S. The union will seek a wage close to the maximum the employer would be willing to pay for this quantity, Wu, at the intersection of the marginal revenue product (MRP) and the marginal factor cost (MFC) curves. The actual wage will be somewhere between these two amounts.*

Figure 14.7 shows the same monopsony situation in a labor market that was shown in Figure 14.3. The employer will seek to pay a wage Wm for a quantity of labor *Lm*. The union will seek Wu, the highest wage the employer would be willing to pay for that quantity of labor. This wage is found on the *MRP* curve. The model of bilateral monopoly does not tell us the wage that will emerge. Whether the final wage will be closer to what the union seeks or closer to what the employer seeks will depend on the bargaining strength of the union and of the employer.

THE DECLINE IN U.S. UNION MEMBERSHIP

The proportion of U.S. workers belonging to unions has declined dramatically since the early 1950s. Economists have offered a number of possible explanations:

- The shift from manufacturing to service industries
- The force of globalization and increased competition from foreign producers
- A reduced desire for unions because of the workplace protection laws now in place
- U.S. legal environment that makes it relatively more difficult for unions to organize workers and expand their membership

Let's discuss each of these four explanations in more detail.

A first possible explanation for the decline in the share of U.S. workers belonging to unions involves the patterns of job growth in the manufacturing and service sectors of the economy shown in Figure 15.4. The U.S. economy had about 15 million manufacturing jobs in 1960. This total rose to 19 million by the late 1970s and then declined to 17 million in 2013. Meanwhile, the number of jobs in service industries and in government combined rose from 35 million in 1960 to over 118 million by 2013, according to the Bureau of Labor Statistics. Because over time unions were stronger in manufacturing than in service industries, the growth in jobs was not happening where the unions were. It is interesting to note that several of the biggest unions in the country are made up of government workers, including the *American Federation of State,County and Municipal Employees* (AFSCME); the *Service Employees International Union*; and the *National Education Association*. The membership of each of these unions is listed in Table 15.2. Outside of government employees, however, unions have not had great success in organizing the service sector.

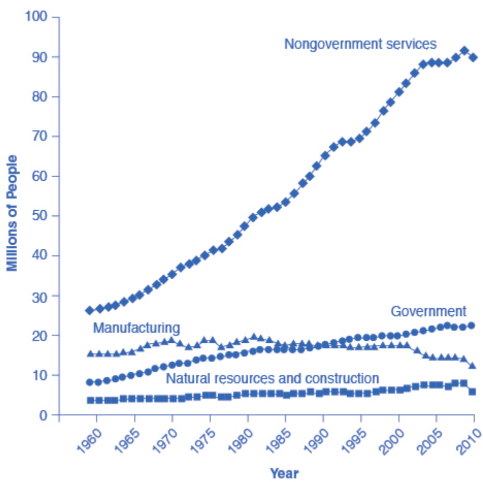


***Figure 15.4.** The Growth of Service Jobs. Jobs in services have increased dramatically in the last few decades. Jobs in government have increased modestly. Jobs in manufacturing have not changed much, although they have trended down in recent years.* Source: U.S. Department of Labor, Bureau of Labor Statistics.

A second explanation for the decline in the share of unionized workers looks at import competition. Starting in the 1960s, U.S. carmakers and steelmakers faced increasing competition from Japanese and European manufacturers. As sales of imported cars and steel rose, the number of jobs in U.S. auto manufacturing fell. This industry is heavily unionized. Not surprisingly, membership in the United Auto Workers, which was 975,000 in 1985, had fallen to roughly 390,000 by 2013. Import competition not only decreases the employment in sectors where unions were once strong, but also decreases the bargaining power of unions in those sectors. However, as we have seen, unions that organize public-sector workers, who are not threatened by import competition, have continued to see growth.

A third possible reason for the decline in the number of union workers is that citizens often call on their elected representatives to pass laws concerning work conditions, overtime, parental leave, regulation of pensions, and other issues. Unions offered strong political support for these laws aimed at protecting workers but, in an ironic twist, the passage of those laws then made many workers feel less need for unions.

These first three possible reasons for the decline of unions are all somewhat plausible, but they have a common problem. Most other developed economies have experienced similar economic and political trends, such as the shift from manufacturing to services, globalization, and increasing government social benefits and regulation of the workplace. Clearly there are cultural differences between countries as to their acceptance of unions in the workplace. The share of the population belonging to unions in other countries is very high compared with the share in the United States. Table 15.4 shows the proportion of workers in a number of the world's high-income economies who belong to unions. The United States is near the bottom, along with France and Spain. The last column shows union coverage, defined as including those workers whose wages are determined by a union negotiation even if the workers do not officially belong to the union. In the United States, union membership is almost identical to union coverage. However, in many countries, the wages of many workers who do not officially belong to a union are still determined by collective bargaining between unions and firms.

Table 15.4. International Comparisons of Union Membership and Coverage in 2012 (Source, CIA World Factbook, retrieved from www.cia.gov)

Country	Union Density: Percentage of Workers Belonging to a Union	Union Coverage: Percentage of Workers Whose Wages Are Determined by Union Bargaining
Austria	37%	99%
France	9%	95%
Germany	26%	63%
Japan	22%	23%
Netherlands	25%	82%
Spain	11.3%	81%
Sweden	82%	92%
United Kingdom	29%	35%
United States	11.3%	12.5%

These international differences in union membership suggest a fourth reason for the decline of union membership in the United States: perhaps U.S. laws are less friendly to the formation of unions than such laws in other countries. The close connection between union membership and a friendly legal environment is apparent in the history of U.S. unions. The great rise in union membership in the 1930s followed the passage of the *National Labor-Management Relations Act* of 1935, which specified that workers had a right to organize unions and that management had to give them a fair chance to do so. The U.S. government strongly encouraged the formation of unions during the early 1940s in the belief that unions would help to coordinate the all-out production efforts needed during World War II. However, after World War II came the passage of the Taft-Hartley Act of 1947, which gave states the power to allow workers to opt out of the union in their workplace if they so desired. This law made the legal climate less encouraging to those seeking to form unions, and union membership levels soon started declining.

The procedures for forming a union differ substantially from country to country. For example, the procedures in the United States and those in Canada are strikingly different. When a group of workers wish to form a union in the United States, they announce this fact and an election date is set when the employees at a firm will vote in a secret ballot on whether to form a union. Supporters of the union lobby for a "yes" vote, and the management of the firm lobbies for a "no" vote—often even hiring outside consultants for assistance in swaying workers to vote "no." In Canada, by contrast, a union is formed when a sufficient proportion of workers (usually about 60%) sign an official card saying that they want a union. There is no separate "election date." The management of Canadian firms is limited by law in its ability to lobby against the union. In addition, though it is illegal to discriminate and fire workers based on their union activity in the United States, the penalties are slight, making this a not so costly way of deterring union activity. In short, forming unions is easier in Canada—and in many other countries—than in the United States.

In summary, union membership in the United States is lower than in many other high-income countries, a difference that may be due to different legal environments and cultural attitudes toward unions.

LINK IT UP

Visit this website to read about recent protests regarding minimum wage for fast food employees.

Unions and the Economy: An Assessment

Where unions operate effectively in otherwise competitive markets, they may reduce economic efficiency. Efforts to increase demand for American workers through restricting imports or to increase demand for skilled workers by restricting opportunities for unskilled workers almost certainly reduce economic efficiency. Artificial restrictions on the supply of labor reduce efficiency as well. In each case, the wage gain will increase the cost of producing a good or service and thus shift its supply curve to the left. Such efforts, if successful, increase the earnings of union members by creating higher prices and smaller quantities for consumers. They may also reduce the profitability of their employers.

Other attempts by unions to raise wages by increasing the demand for their members are not likely to create inefficiency. For example, union efforts to increase worker productivity or to encourage consumers to buy products made by union members do not reduce economic efficiency.

In the case of bilateral monopoly, the amount of labor employed is restricted by the monopsony firm to a quantity that falls short of the efficient level. In effect, the efficiency damage has already been done. The labor union seeks merely to offset the monopsony firm's ability to restrict the wage.

Are unions successful in their primary goal of increasing wages? An examination of the impact on wages paid by firms that faced organizing drives by unions between 1984 and 1999 found virtually no change in wages attributable to union organizing efforts. The study examined firms in which unions had either barely won or had barely lost the election. It found that unions that had eked out victories had gone on to organize workers but had had no significant impact on wages or on productivity. Other evidence, however, suggests that unions do tend to raise wages for their members. Controlling for other factors that affect wages, over the period 1973 to 2002, unions appear to have increased wages by about 17% on average. Part of the explanation of this finding is that unions have had the most success in organizing in the public sector, where union pressure for higher wages is most likely to be successful.

Other Suppliers and Monopoly Power

Just as workers can unionize to gain a degree of monopoly power in the marketplace, so other suppliers can organize with a similar goal. Two of the most important types of organizations aimed at garnering market power are professional associations and producers' cooperatives.

Professional Associations

Professional people generally belong to organizations that represent their interests. For example, physicians in the United States belong to the American Medical Association (AMA), and lawyers belong to the American Bar Association (ABA). Both organizations work vigorously to advance the economic interests of their members.

Professional organizations often lobby for legislation that protects their members. They may seek to restrict competition by limiting the number of individuals who can be licensed to practice a particular profession. The AMA has been very successful in limiting the number of physicians, thus maintaining higher salaries than would otherwise exist. The ABA has fought legal reforms aimed at limiting awards to plaintiffs who win damage suits; such reforms would be likely to reduce the incomes of lawyers.

Producers' Cooperatives

Independent producers sometimes band together into a cooperative for the purpose of selling their products. The cooperative sets the price and assigns production quotas to individual firms. In effect, a cooperative acts as a legal cartel.

Because they violate the provisions of laws that outlaw such arrangements in most industries, producers' cooperatives must be authorized by Congress. Farmers have sometimes been given such rights when they are confronted by monopsony buyers. For example, Congress granted dairy farmers the right to form cooperatives in the 1920s because they faced monopsony buyers. High transportation costs for fresh milk, together with economies of scale in processing milk, generally left only one dairy processor to buy raw milk from dairy farmers in a particular area. By forming a cooperative, farmers could counter the monopsony power of a processor with monopoly power of their own, creating a bilateral monopoly.

Today, with much lower transportation costs, dairy farmers can deal with a national market so that processors no longer have monopsony power. But dairy farmers continue to have the right to form cooperatives. As we have seen in an earlier module, dairy farmers also enjoy protection from federal programs that are designed to keep dairy prices high.

KEY TAKEAWAYS

- A firm that has monopoly power in the supply of a factor makes choices in the same manner as any monopoly firm; it maximizes profit by selecting a level of output at which marginal revenue equals marginal cost and selling that output at a price determined by the demand curve.

- Unions have traditionally sought to raise wages and to improve working conditions by exerting market power over the supply of labor.
- In bilateral monopoly, a monopsony buyer faces a monopoly seller. Prices in the model are indeterminate.
- Professional associations often seek market power through their influence on government policy.
- Producers' cooperatives, a form of legal cartel, have been organized in some agricultural markets in an effort to offset the perceived monopsony power of some buyers of agricultural products.

Case in Point: Unions and the Airline Industry

Unions represent 60% of the non-managerial employees of U.S. airlines. And labor costs make up one-third of airline costs. All employees have a stake in the success of the firms for which they work. That is certainly the case for the major unions representing airline employees. Both union leaders and airline management have much to gain from a relationship that benefits both employees and the airlines that employ them.

That sort of relationship has not always existed. In 1981, for example, Continental Airlines hired Frank Lorenzo, an airline entrepreneur, to run Continental. The airline had lost money the previous three years. Mr. Lorenzo promptly abrogated Continental's contracts with employees, and told them that they could go back to work but only at sharply reduced wages. Continental's pilots, flight attendants, and ground crews declared strikes against the airline. The airline was able to break the strike by hiring replacement employees. Even so, Continental declared bankruptcy in 1983. Mr. Lorenzo told striking employees that they could return to work, but they could do so only by agreeing to work at half their previous wage. Continental's strategy of union suppression achieved reductions in wage costs, but those savings had a cost as well. A demoralized labor force produced dramatic reductions in the quality of service, and Continental was back in bankruptcy in 1991. In 1986, 6,000 members of the International Federation of Flight Attendants (IFFA) declared a strike against TWA. The airline followed a strategy similar to Continental's and was able to break the strike by hiring replacement employees. After 10 weeks, the IFFA declared an unconditional end to the strike and sought to have its members rehired. It was not until three and a half years later that all 6,000 got their jobs back. Ultimately, TWA went out of business.

Not all airlines have had the same unhappy relationships with unions. Southwest Airlines, which started as a nonunion airline, now operates with a largely unionized labor force. It has continued its strategy of paying high wages and including employees in management decisions. The result has been one of the highest profit margins in the industry together with high productivity of both workers and aircraft.

Continental has also emerged as a "different" airline. More recently, it hired a new Chief Executive Officer who quickly returned the airline to profitability and established a new workplace culture in which employees were given a role in managerial decisions and were hired based in part on their teamwork skills. Continental has been able to shed its old reputation as a union suppressor and has established itself as an airline that works well with unions and has a minimal degree of conflict.

Another approach to dealing with airline employees has been to include them in ownership. United employees, for example, own the airline. Other airlines in which employees have had a substantial ownership role include Western Airlines, Eastern Airlines, Northwest, Delta Airlines, and United. In each case, employees exchanged equity for wage concessions. Each of these airlines implemented an Employee Stock Ownership Plan (ESOP). In each case, the program began with great optimism on the part of management and labor, but in most cases, conflicts between workers and their employers quickly emerged. Western and Eastern abandoned their ESOP programs after two years. While nearly all of the ESOP arrangements initially increased profits, none of them was accompanied by any structural change in the labor-management relationship. Ultimately, all of these plans generated disappointing results. Clearly, the mere creation of a system in which employees own a share of the airline is not sufficient; changes in the structure of the labor-management relations must occur as well. Some airlines have managed to prosper in a difficult economic world. The key to success seems to lie in the establishment of workplace culture that rewards good teamwork and efforts to enhance productivity. Airlines such as Southwest and the "new" Continental demonstrate that an airline can work effectively with unions, pay high wages, and still be profitable.

313. Simulation: Income Distribution

Try It

Play the simulation below multiple times to see how different choices lead to different outcomes. All simulations allow unlimited attempts so that you can gain experience applying the concepts.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=346

314. Self Check: Competition and Wages

Check Your Understanding

Answer the question(s) below to see how well you understand the topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the four Readings in this section.

Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/700

315. Outcome: Measuring and Understanding the Distribution of Income

What you'll learn to do: use the Lorenz Curve to analyze the distribution of income and wealth

In this outcome, you will use the Lorenz Curve to examine the unequal distribution of income in the United States and take a closer look at the poverty trap and what the government does to reduce income inequality.

LEARNING ACTIVITIES

The learning activities for this section include the following:

- Reading: Poverty and Economic Inequality
- Reading: The Poverty Trap
- Reading: The Safety Net
- Reading: Explaining Inequality
- Reading: Measuring Income Inequality
- Reading: Government Policies to Reduce Income Inequality
- Self Check: Measuring and Understanding the Distribution of Income

Take time to review and reflect on each of these activities in order to improve your performance on the assessment for this section.

316. Reading: Poverty and Economic Inequality

This image shows hundreds of protestors during Occupy Wall Street.

Figure 14.1. *Occupying Wall Street. On September 17, 2011, Occupy Wall Street began in New York City's Wall Street financial district. (Credit: modification of work by David Shankbone/Flickr Creative Commons)*

OCCUPY WALL STREET

In September 2011, a group of protesters gathered in Zuccotti Park in New York City to decry what they perceived as increasing social and economic inequality in the United States. Calling their protest "Occupy Wall Street," they argued that the concentration of wealth among the richest 1% in the United States was both economically unsustainable and inequitable, and needed to be changed. The protest then spread to other major cities, and the Occupy movement was born.

Why were people so upset? How much wealth is concentrated among the top 1% in our society? How did they acquire so much wealth? These are very real, very important questions in the United States now, and this module on poverty and economic inequality will help us address the causes behind this sentiment.

Introduction to Poverty and Economic Inequality

The labor markets that determine what workers are paid do not take into account how much income a family needs for food, shelter, clothing, and health care. Market forces do not worry about what happens to families when a major local employer goes out of business. Market forces do not take time to contemplate whether those who are earning higher incomes should pay an even higher share of taxes.

However, labor markets do create considerable inequalities of income. In 2012, the median American family income was $62,241 (the median is the level where half of all families had more than that level and half had less). According to the U.S. Census Bureau, almost nine million U.S. families were classified by the federal government as being below the poverty line in that year. Think about a family of three—perhaps a single mother with two children—attempting to pay for the basics of life on perhaps $17,916 per year. After paying for rent, healthcare, clothing, and transportation, such a family might have $6,000 to spend on food. Spread over 365 days, the food budget for the entire family would be about $17 per day. To put this in perspective, most cities have restaurants where $17 will buy you an appetizer for one.

This module explores how the U.S. government defines poverty, the balance between assisting the poor without discouraging work, and how federal antipoverty programs work. It also discusses income inequality—how economists measure inequality, why inequality has changed in recent decades, the range of possible government policies to reduce inequality, and the danger of a tradeoff that too great a reduction in inequality may reduce incentives for producing output.

Drawing the Poverty Line

Comparisons of high and low incomes raise two different issues: economic inequality and *poverty*. Poverty is measured by the number of people who fall below a certain level of income—called the ***poverty line***—that defines the income needed for a basic standard of living. ***Incomeinequality*** compares the share of the total income (or wealth) in society that is received by different groups; for example, comparing the share of income received by the top 10% to the share of income received by the bottom 10%.

In the United States, the official definition of the poverty line traces back to a single person: *Mollie Orshansky*. In 1963, Orshansky, who was working for the Social Security Administration, published an article called "Children of the Poor" in a highly useful and dry-as-dust publication called the *Social Security Bulletin*. Orshansky's idea was to define a poverty line based on the cost of a healthy diet.

Her previous job had been at the U.S. Department of Agriculture, where she had worked in an agency called the Bureau of Home Economics and Human Nutrition. One task of this bureau had been to calculate how much it would cost to feed a nutritionally adequate diet to a family. Orshansky found that the average family spent one-third of its income on food. She then proposed that the poverty line be the amount needed to buy a nutritionally adequate diet, given the size of the family, multiplied by three.

The current U.S. poverty line is essentially the same as the Orshansky poverty line, although the dollar amounts are adjusted each year to represent the same buying power over time. The U.S. poverty line in 2012 ranged from $11,720 for a single individual to $23,492 for a household of four people.

Figure 14.2 shows the U.S. ***poverty rate*** over time; that is, the percentage of the population below the poverty line in any given year.

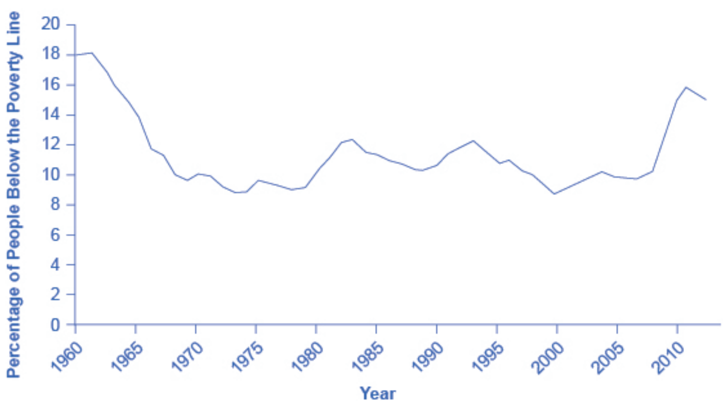


Figure 14.2. *The U.S. Poverty Rate since 1960. The poverty rate fell dramatically during the 1960s, rose in the early 1980s and early 1990s, and, after declining in the 1990s through mid-2000s, rose to 15.9% in 2011, which is close to the 1960 levels. In 2012, the poverty dropped slightly to 15.0%. (Source: U.S. Census Bureau)*

The poverty rate declined through the 1960s, rose in the early 1980s and early 1990s, but seems to have been slightly lower since the mid-1990s. However, in no year in the last four decades has the poverty rate been less than 11% of the U.S. population—that is, at best about one American in nine is below the poverty line. In recent years, the poverty rate appears to have peaked at 15.9% in 2011 before dropping to 15.0% in 2012. Table 14.1 compares poverty rates for different groups in 2011.

Table 14.1 Poverty Rates by Group, 2011

Group	Poverty Rate
Females	16.3%
Males	13.6%
White	13.0%
Black	27.6%
Hispanic	25.3%
Under age 18	21.9%
Ages 18–24	20.6%
Ages 25–34	15.9%
Ages 35–44	12.2%
Ages 45–54	10.9%
Ages 55–59	10.7%
Ages 60–64	10.8%
Ages 65 and older	8.7%

As you will see when we delve further into these numbers, poverty rates are relatively low for whites, for the elderly, for the well-educated, and for male-headed households. Poverty rates for females, Hispanics, and African Americans are much higher than for whites. While Hispanics and African Americans have a higher percentage of individuals living in poverty than others, most people in the United States living below the poverty line are white.

The concept of a poverty line raises many tricky questions. In a vast country like the United States, should there be a national poverty line? After all, according to the Federal Register, the median household income for a family of four was $102,552 in New Jersey and $57,132 in Mississippi in 2013, and prices of some basic goods like housing are quite different between states. The poverty line is based on cash income, which means it does not take into account government programs that provide assistance to the poor in a non-cash form, like *Medicaid* (health care for low-income individuals and families) and food aid. Also, low-income families can qualify for federal housing assistance. (These and other government aid programs will be discussed in detail later in this module.)

Should the poverty line be adjusted to take the value of such programs into account? Many economists and policymakers wonder whether the concept of what poverty means in the twenty-first century should be rethought. The following section explains the poverty lines set by the World Bank for low-income countries around the world.

Link It Up

Visit this website for more information on U.S. poverty.

HOW IS POVERTY

MEASURED IN LOW-INCOME COUNTRIES?

The World Bank sets two poverty lines for low-income countries around the world. One poverty line is set at an income of $1.25/day per person; the other is at $2/day. By comparison, the U.S. 2011 poverty line of $17,916 annually for a family of three works out to $16.37 per person per day.

Clearly, many people around the world are far poorer than Americans, as Table 14.2 shows. China and India both have more than a billion people; Nigeria is the most populous country in Africa; and Egypt is the most populous country in the Middle East. In all four of those countries, in the mid-2000s, a substantial share of the population subsisted on less than $2/day. Indeed, about half the world lives on less than $2.50 a day, and 80 percent of the world lives on less than $10 per day. (Of course, the cost of food, clothing, and shelter in those countries can be very different from those costs in the United States, so the $2 and $2.50 figures may mean greater purchasing power than they would in the United States.)

Table 14.2. Poverty Lines for Low-Income Countries, mid-2000s (Source: http://data.worldbank.org/indicator/SI.POV.DDAY)

Country	Share of Population below $1.25/Day	Share of Population below $2.00/Day
Brazil (in 2009)	6.1%	10.8%
China (in 2009)	11.8%	27.2%
Egypt (in 2008)	1.7%	15.4%
India (in 2010)	32.7%	68.8%
Mexico (in 2010)	0.7%	4.5%
Nigeria (in 2010)	68.0%	84.5%

Any poverty line will be somewhat arbitrary, and it is useful to have a poverty line whose basic definition does not change much over time. If Congress voted every few years to redefine what poverty means, then it would be difficult to compare rates over time. After all, would a lower poverty rate mean that the definition had been changed, or that people were actually better off? Government statisticians at the U.S. Census Bureau have ongoing research programs to address questions like these.

317. Reading: The Poverty Trap

The Poverty Trap

Can you give people too much help, or the wrong kind of help? When people are provided with food, shelter, healthcare, income, and other necessities, assistance may reduce their incentive to work. Consider a program to fight poverty that works in this reasonable-sounding manner: the government provides assistance to the poor, but as the poor earn income to support themselves, the government reduces the level of assistance it provides. With such a program, every time a poor person earns $100, the person loses $100 in government support. As a result, the person experiences no net gain for working. Economists call this problem the *poverty trap*.

Consider the situation faced by a single-parent family. A single mother (earning $8 an hour) with two children, as illustrated in Figure 14.3. First, consider the *labor-leisure budget constraint* faced by this family in a situation without government assistance. On the horizontal axis is hours of leisure (or time spent with family responsibilities) increasing in quantity from right to left. Also on the horizontal axis is the number of hours at paid work, going from zero hours on the right to the maximum of 2,500 hours on the left. On the vertical axis is the amount of income per year rising from low to higher amounts of income. The budget constraint line shows that at zero hours of leisure and 2,500 hours of work, the maximum amount of income is $20,000 ($8 × 2,500 hours). At the other extreme of the budget constraint line, an individual would work zero hours, earn zero income, but enjoy 2,500 hours of leisure. At point A on the *budget constraint line*, by working 40 hours a week, 50 weeks a year, the utility-maximizing choice is to work a total of 2,000 hours per year and earn $16,000.

Now suppose that a government antipoverty program guarantees every family with a single mother and two children $18,000 in income. This is represented on the graph by a horizontal line at $18,000. With this program, each time the mother earns $1,000, the government will deduct $1,000 of its support. Table 14.3 shows what will happen at each combination of work and government support.

The graph shows a downward sloping line that begins at $20,000 on the y-axis and ends at 2,500 on the x-axis. A horizontal line extends from $18,000 on the y-axis. A dashed plum line extends from $16,000 on the y-axis and intersects with the vertical line extending from 500 on the x-axis at point A. Beneath the x-axis is an arrow pointing to the right indicating leisure (hours) and an arrow pointing to the left indicating labor (hours).

Figure 14.3. *The Poverty Trap in Action. The original choice is 500 hours of leisure, 2,000 hours of work at point A, and income of $16,000. With a guaranteed income of $18,000, this family would receive $18,000 whether it provides zero hours of work or 2,000 hours of work. Only if the family provides, say, 2,300 hours of work does its income rise above the guaranteed level of $18,000—and even then, the marginal gain to income from working many hours is small.*

Table 14.3. Total Income at Various Combinations of Work and Support

Amount Worked (hours)	Total Earnings	Government Support	Total Income
0	0	$18,000	$18,000
500	$4,000	$14,000	$18,000
1,000	$8,000	$10,000	$18,000
1,500	$12,000	$6,000	$18,000
2,000	$16,000	$2,000	$18,000
2,500	$20,000	0	$20,000

The new budget line, with the antipoverty program in place, is the horizontal and heavy line that is flat at $18,000. If the mother does not work at all, she receives $18,000, all from the government. If she works full time, giving up 40 hours per week with her children, she still ends up with $18,000 at the end of the year. Only if she works 2,300 hours in the year—which is an average of 44 hours per week for 50 weeks a year—does household income rise to $18,400. Even in this case, all of her year's work means that household income rises by only $400 over the income she would receive if she did not work at all. She would need to work 50 hours a week to reach $20,000.

Indeed, the poverty trap is even stronger than this simplified example shows, because a working mother will have extra expenses like clothing, transportation, and child care that a nonworking mother will not face, making the economic gains from working even smaller. Moreover, those who do not work fail to build up job experience and contacts, which makes working in the future even less likely.

The bite of the poverty trap can be reduced by designing an antipoverty program so that, instead of reducing government payments by $1 for every $1 earned, payments are reduced by some smaller amount instead. The bite of the poverty trap can also be reduced by imposing requirements for work as a condition of receiving benefits and setting a time limit on benefits.

Figure 14.4 illustrates a government program that guarantees

$18,000 in income, even for those who do not work at all, but then reduces this amount by 50 cents for each $1 earned. The new, higher budget line in Figure 14.4 shows that, with this program, additional hours of work will bring some economic gain. Because of the reduction in government income when an individual works, an individual earning $8.00 will really net only $4.00 per hour. The vertical intercept of this higher budget constraint line is at $28,000 ($18,000 + 2,500 hours × $4.00 = $28,000). The horizontal intercept is at the point on the graph where $18,000 and 2500 hours of leisure is set. Table 14.4 shows the total income differences with various choices of labor and leisure.

However, this type of program raises other issues. First, even if it does not eliminate the incentive to work by reducing government payments by $1 for every $1 earned, enacting such a program may still reduce the incentive to work. At least some people who would be working 2,000 hours each year without this program might decide to work fewer hours but still end up with more income—that is, their choice on the new budget line would be like S, above and to the right of the original choice P. Of course, others may choose a point like R, which involves the same amount of work as P, or even a point to the left of R that involves more work.

The second major issue is that when the government phases out its support payments more slowly, the antipoverty program costs more money. Still, it may be preferable in the long run to spend more money on a program that retains a greater incentive to work, rather than spending less money on a program that nearly eliminates any gains from working.

The graph shows a downward sloping line that extends from $28,000 on the y-axis to $18,000 on the y-axis (from 0 to 2,500 on the x-axis). Two points R and S appear on the line. Another line starts at (0, $20,000) and ends at (2,500, 0). A dashed plum line extends horizontally from $18,000 on the y-axis and meets with the vertical line extending from 2,500 on the x-axis. Another dashed plum line extends from $16,000 on the y-axis and intersects with the vertical line extending from 500 on the x-axis at point P. Beneath the x-axis is an arrow pointing to the right indicating leisure (hours) and an arrow pointing to the left indicating labor (hours).

Figure 14.4. *Loosening the Poverty Trap: Reducing Government Assistance by 50 Cents for Every $1 Earned. On the original labor-leisure opportunity set, the lower budget set shown by the smaller dashed line in the figure, the preferred choice P is 500 hours of leisure and $16,000 of income. Then, the government created an antipoverty program that guarantees $18,000 in income even to those who work zero hours, shown by the larger dashed line. In addition, every $1 earned means phasing out 50 cents of benefits. This program leads to the higher budget set shown in the diagram. The hope is that this program will provide incentives to work the same or more hours, despite receiving income assistance. However, it is possible that the recipients will choose a point on the new budget set like S, with less work, more leisure, and greater income, or a point like R, with the same work and greater income.*

Table 14.4. The Labor-Leisure Tradeoff with Assistance Reduced by 50 Cents for Every Dollar Earned

Amount Worked (hours)	Total Earnings	Government Support	Total Income
0	0	$18,000	$18,000
500	$4,000	$16,000	$20,000
1,000	$8,000	$14,000	$22,000
1,500	$12,000	$12,000	$24,000
2,000	$16,000	$10,000	$26,000
2,500	$20,000	$8,000	$28,000

The next module will consider a variety of government support programs focused specifically on the poor, including welfare, SNAP (food supplement), Medicaid, and the earned income tax credit (EITC). Although these programs vary from state to state, it is generally a true statement that in many states from the 1960s into the 1980s, if poor people worked, their level of income barely rose—or did not rise at all—after the reduction in government support payments was factored in.

TRY IT

CALCULATING A BUDGET CONSTRAINT LINE

Jason earns $9.00 an hour, and a government antipoverty program provides a floor of $10,000 guaranteed income. The government reduces government support by $0.50 for each $1.00 earned. What are the horizontal and vertical intercepts of the budget constraint line? Assume the maximum hours for work or leisure is 2,500 hours.

Step 1. Determine the amount of the government guaranteed income. In this case, it is $10,000.

Step 2. Plot that guaranteed income as a horizontal line on the budget constraint line.

Step 3. Determine what Jason earns if he has no income and enjoys 2,500 hours of leisure. In this case, he will receive the guaranteed $10,000 (the horizontal intercept).

Step 4. Calculate how much Jason's salary will be reduced by due to the reduction in government income. In Jason's case, it will be reduced by one half. He will, in effect, net only $4.50 an hour.

Step 5. If Jason works 1,000 hours, at a maximum what income will Jason receive? Jason will get the government assistance of $10,000. He will net only $4.50 for every hour he chooses to work. If he works 1,000 hours at $4.50, his earned income is $4,500 plus the government income of $10,000. Thus the total maximum income (the vertical intercept) is $10,000 + $4,500 = $14,500.

318. Reading: The Safety Net

The Safety Net

The U.S. government has implemented a number of programs to assist those below the poverty line and those who have incomes just above the poverty line, who are referred to as the *near-poor*. Such programs are called the *safety net*, in recognition of the fact that they offer some protection for those who find themselves without jobs or income.

Temporary Assistance for Needy Families

From the Great Depression of the 1930s until 1996, the United States' most visible antipoverty program was *Aid to Families with Dependent Children* (AFDC), which provided cash payments to mothers with children who were below the poverty line. This program was often just called "welfare." In 1996, Congress passed and President Bill Clinton signed into law the Personal Responsibility and Work Opportunity Reconciliation Act, more commonly called the "welfare reform act." The new law replaced AFDC with *Temporary Assistance for Needy Families* (TANF).

LINK IT UP

Visit this website to watch a video of President Bill Clinton's Welfare Reform speech.

TANF brought several dramatic changes in how welfare operated.

Under the old AFDC program, states set the level of welfare benefits that they would pay to the poor, and the federal government guaranteed it would chip in some of the money as well. The federal government's welfare spending would rise or fall depending on the number of poor people, and on how each state set its own welfare contribution.

Under TANF, however, the federal government gives a fixed amount of money to each state. The state can then use the money for almost any program with an antipoverty component: for example, the state might use the money to give cash to poor families, or to reduce teenage pregnancy, or even to raise the high school graduation rate. However, the federal government imposed two key requirements. First, if states are to keep receiving the TANF grants, they must impose work requirements so that most of those receiving TANF benefits are working (or attending school). Second, no one can receive TANF benefits with federal money for more than a total of five years over his or her lifetime. The old AFDC program had no such work requirements or time limits.

TANF attempts to avoid the poverty trap by requiring that welfare recipients work and by limiting the length of time they can receive benefits. In its first few years, the program was quite successful. The number of families receiving payments in 1995, the last year of AFDC, was 4.8 million. By 2012, according to the Congressional Research Service, the average number of families receiving payments under TANF was 1.8 million—a decline of more than half.

TANF benefits to poor families vary considerably across states. For example, again according to the Congressional Research Service, in 2011 the highest monthly payment in Alaska to a single mother with two children was $923, while in Mississippi the highest monthly payment to that family was $170. These payments reflect differences in states' cost of living. Total spending on TANF was approximately $16.6 billion in 1997. As of 2012, spending was at $12 billion, an almost 28% decrease, split about evenly between the federal and state governments. When you take into account the effects of inflation, the decline is even greater. Moreover, there seemed little evidence that poor families were suffering a reduced standard of living as a result of TANF—although, on the other side, there was not much evidence that poor families had greatly improved their total levels of income, either.

The Earned Income Tax Credit (EITC)

The *earned income tax credit* (EITC), first passed in 1975, is a method of assisting the working poor through the tax system. The EITC is the second largest assistance program for low-income groups (after SNAP, described below), and projections for 2013 expected 26 million households to take advantage of it at an estimated cost of $50 billion. In 2013, for example, a single parent with two children would have received a tax credit of $5,372 up to an income level of $17,530. The amount of the tax break increases with the amount of income earned, up to a point. The earned income tax credit has often been popular with both economists and the general public because of the way it effectively increases the payment received for work.

What about the danger of the poverty trap that every additional $1 earned will reduce government support payments by close to $1? To minimize this problem, the earned income tax credit is phased out slowly. According to the Tax Policy Center, for a single-parent family with two children in 2013, the credit is not reduced at all (but neither is it increased) as earnings rise from $13,430 to $17,530. Then, for every $1 earned above $17,530, the amount received from the credit is reduced by 21.06 cents, until the credit phases out completely at an income level of $46,227.

Figure 14.5 illustrates that the earned income tax credits, child tax credits, and the TANF program all cost the federal government money—either in direct outlays or in loss of tax revenues. CTC stands for the government tax cuts for the child tax credit.

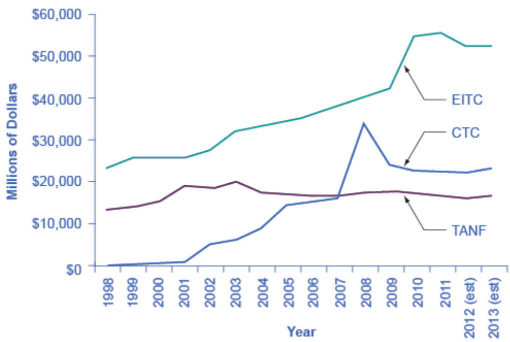


Figure 14.5. *Real Federal Spending on CTC, EITC, and TANF, 1975–2013. EITC increased from more than $20 billion in 2000 to over an estimated $50 billion by 2013, far exceeding estimated 2013 outlays in the CTC (Child Tax Credits) and TANF of over $20 billion and $10 billion, respectively. (Source: Office of Management and Budget)*

In recent years, the EITC has become the single most expensive government program for providing income assistance to the poor and near-poor, costing about $60 billion in 2012. In that year, the EITC provided benefits to about 27 million families and individuals and, on average, is worth about $2,296 per family (with children), according to the Tax Policy Center. One reason that the TANF law worked as well as it did is that the EITC was greatly expanded in the late 1980s and again in the early 1990s, which increased the returns to work for low-income Americans.

Supplemental Nutrition Assistance Program (SNAP)

Often called "food stamps," *Supplemental Nutrition Assistance Program* (SNAP) is a federally funded program, started in 1964, in which each month poor people receive a card like a debit card that they can use to buy food. The amount of food aid for which a household is eligible varies by income, number of children, and other factors but, in general, households are expected to spend about 30% of their own net income on food, and if 30% of their net income is not enough to purchase a nutritionally adequate diet, then those households are eligible for SNAP.

SNAP can contribute to the poverty trap. For every $100 earned, the government assumes that a family can spend $30 more for food, and thus reduces its eligibility for food aid by $30. This decreased benefit is not a complete disincentive to work—but combined with how other programs reduce benefits as income increases, it adds to the problem. SNAP, however, does try to address the poverty trap with its own set of work requirements and time limits.

Why give debit cards and not just cash? Part of the political support for SNAP comes from a belief that since the cards must be spent on food, they cannot be "wasted" on other forms of consumption. From an economic point of view, however, the belief that cards must increase spending on food seems wrong-headed. After all, say that a poor family is spending $2,500 per year on food, and then it starts receiving $1,000 per year in SNAP aid. The family might react by spending $3,500 per year on food (income plus aid), or it might react by continuing to spend $2,500 per year on food, but use the $1,000 in food aid to free up $1,000 that can now be spent on other goods. So it is reasonable to think of SNAP cards as an alternative method, along with TANF and the earned income tax credit, of transferring income to the working poor.

Indeed, anyone eligible for TANF is also eligible for SNAP, although states can expand eligibility for food aid if they wish to do so. In some states, where TANF welfare spending is relatively low, a poor family may receive more in support from SNAP than from TANF. In 2012, about 46.6 million people received food aid at an annual cost of about $74.6 billion, with an average monthly benefit of about $287 per person per month. SNAP participation increased by 70% between 2007 and 2011, from 26.6 million participants to 45 million. According to the Congressional Budget Office, this dramatic rise in participation was caused by the *Great Recession* of 2008–2009 and rising food prices.

The federal government deploys a range of income security programs that are funded through departments such as Health and Human Services, Agriculture, and Housing and Urban Development (HUD) (see Figure 14.6). According to the Office of Management and Budget, collectively, these three departments provided an estimated $62 billion of aid through programs such as supplemental feeding programs for women and children, subsidized housing, and energy assistance. The federal government also transfers funds to individual states through special grant programs.

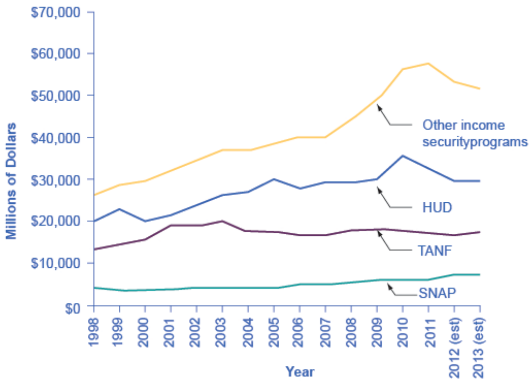


Figure 14.6. Expenditure Comparison of TANF, SNAP, HUD, and Other Income Security Programs, 1988–2013 (est.). Total expenditures on income security continued to rise between 1988 and 2010, while payments for TANF have increased from $13 billion in 1998 to an estimated $17.3 billion in 2013. SNAP has seen relatively small increments. These two programs comprise a relatively small portion of the estimated $106 billion dedicated to income security in 2013. Note that other programs and housing programs increased dramatically during the

2008 and 2010 time periods. (Source: Table 12.3 Section 600 Income Security, http://www.whitehouse.gov/sites/default/files/omb/budget/fy2013/assets/hist.pdf)

The safety net includes a number of other programs: government-subsidized school lunches and breakfasts for children from low-income families; the *Special Supplemental Food Program for Women, Infants and Children* (WIC), which provides food assistance for pregnant women and newborns; the Low Income Home Energy Assistance Program, which provides help with home heating bills; housing assistance, which helps pay the rent; and Supplemental Security Income, which provides cash support for the disabled and the elderly poor.

Medicaid

Medicaid was created by Congress in 1965 and is a joint health insurance program entered into by both the states and the federal government. The federal government helps fund Medicaid, but each state is responsible for administering the program, determining the level of benefits, and determining eligibility. It provides medical insurance for certain low-income people, including those below the poverty line, with a focus on families with children, the elderly, and the disabled. About one-third of Medicaid spending is for low-income mothers with children. While an increasing share of the program funding in recent years has gone to pay for nursing home costs for the elderly poor. The program ensures that a basic level of benefits is provided to Medicaid participants, but because each state sets eligibility requirements and provides varying levels of service, the program differs from state to state.

In the past, a common problem has been that many low-paying jobs pay enough to a breadwinner so that a family could lose its eligibility for Medicaid, yet the job does not offer health insurance benefits. A poor parent considering such a job might choose not to work rather than lose health insurance for his or her children. In this way, health insurance can become a part of the poverty trap. Many states recognized this problem in the 1980s and 1990s and expanded their Medicaid coverage to include not just the poor, but the near-poor earning up to 135% or even 185% of the poverty line. Some states also guaranteed that children would not lose coverage if their parents worked.

These expanded guarantees cost the government money, of course, but they also helped to encourage those on welfare to enter the labor force. As of 2012, approximately 67 million people participated in Medicaid. Of those enrolled, almost half are children. Healthcare expenditures, however, are highest for the elderly population, which comprises approximately 25% of participants. As Figure 14.7 (a) indicates, the largest number of households that enroll in Medicaid are those with children. Lower-income adults are the next largest group enrolled in Medicaid at 28%. The blind and disabled are 16% of those enrolled, and seniors are 9% of those enrolled. Figure 14.7 (b) shows how much actual Medicaid dollars are spent for each group. Out of total Medicaid spending, more is spent on seniors (20%) and the blind and disabled (44%). So, 64% of all Medicaid spending goes to seniors, the blind, and disabled. Children receive 21% of all Medicaid spending, followed by adults at 15%.

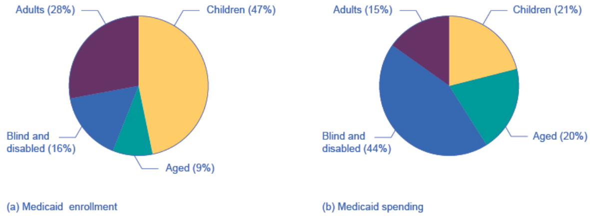


Figure 14.7. *Medicaid Enrollment and Spending. Part (a) shows the Medicaid enrollment by different populations, with children comprising the largest percentage at 47%, followed by adults at 28%, and the blind and disabled at 16%. Part (b) shows that Medicaid spending is principally for the blind and disabled, followed by the elderly. Although children are the largest population covered by Medicaid, expenditures on children are only at 21%.*

319. Reading: Explaining Inequality

Explaining Inequality

Everyone agrees that the distribution of income in the United States generally became more equal during the first two decades after World War II and that it has become more unequal since 1968. While some people conclude that this increase in inequality suggests the latter period was unfair, others want to know why the distribution changed. We shall examine some of the explanations.

Family Structure

Clearly, an important source of rising inequality since 1968 has been the sharp increase in the number of families headed by women. In 2010, the median income of families headed by married couples was 2.5 times that of families headed by women without a spouse. The percentage of families headed by women with no spouse present has nearly doubled since 1968 and is thus contributing to increased inequality across households.

Technological and Managerial Change

Technological change has affected the demand for labor. One of the most dramatic changes since the late 1970s has been an increase in the demand for skilled labor and a reduction in the demand for unskilled labor.

The result has been an increase in the gap between the wages of skilled and unskilled workers. That has produced a widening gap between college- and high-school-trained workers.

Technological change has meant the integration of computers into virtually every aspect of production. And that has increased the demand for workers with the knowledge to put new methods to work—and to adapt to the even more dramatic changes in production likely to come. At the same time, the demand for workers who do not have that knowledge has fallen.

Along with new technologies that require greater technical expertise, firms are adopting new management styles that require stronger communication skills. The use of production teams, for example, shifts decision-making authority to small groups of assembly-line workers. That means those workers need more than the manual dexterity that was required of them in the past. They need strong communication skills. They must write effectively, speak effectively, and interact effectively with other workers. Workers who cannot do so simply are not in demand to the degree they once were.

The "intellectual wage gap" seems likely to widen as we move even further into the twenty-first century. That is likely to lead to an even higher degree of inequality and to pose a challenge to public policy for decades to come. Increasing education and training could lead to reductions in inequality. Indeed, individuals seem to have already begun to respond to this changing market situation, since the percentage who graduate from high school and college is rising.

Tax Policy

Did tax policy contribute to rising inequality over the past four decades? The tax changes most often cited in the fairness debate are the Reagan tax cuts introduced in 1981 and the Bush tax cuts introduced in 2001, 2002, and 2003.

An analysis of the Bush tax cuts by the Tax Foundation combines the three Bush tax cuts and assumes they occurred in 2003. Table 19.1 "Income Tax Liability Before and After the Bush Tax Cuts" gives the share of total income tax liability for each quintile before and after the Bush tax cuts. It also gives the share of the Bush tax cuts received by each quintile.

Table 19.1 Income Tax Liability Before and After the Bush Tax Cuts

Quintile	Share of income tax liability before tax cuts	Share of income tax liability after tax cuts	Share of total tax relief
First quintile	0.5%	0.3%	1.2%
Second quintile	2.3%	1.9%	4.2%
Third quintile	5.9%	5.2%	9.4%
Fourth quintile	12.6%	11.6%	17.5%
Top quintile	78.7%	81.0%	67.7%

The share of total tax relief received by the first four quintiles was modest, while those in the top quintile received more than two-thirds of the total benefits of the three tax cuts. However, the share of income taxes paid by each of the first four quintiles fell as a result of the tax cuts, while the share paid by the top quintile rose.

Tax cuts under George W. Bush were widely criticized as being tilted unfairly toward the rich. And certainly, Table 19.1 "Income Tax Liability Before and After the Bush Tax Cuts" shows that the share of total tax relief received by the first four quintiles was modest, while those in the top quintile garnered more than two-thirds of the total benefits of the three tax cuts. Looking at the second and third columns of the table, however, gives a different perspective. The share of income taxes paid by each of the first four quintiles fell as a result of the tax cuts, while the share paid by the top quintile rose. Further, we see that each of the first four quintiles paid a very small share of income taxes before and after the tax cuts, while those in the top quintile ended up shouldering more than 80% of the total income tax burden. We saw in Figure 19.1 "The Distribution of U.S. Income, 1968 and 2010" that those in the top quintile received just over half of total income. After the Bush tax cuts, they paid 81% of income taxes. Others are quick to point out that those same tax cuts were accompanied by reductions in expenditures for some social service programs designed to help lower income families. Still others point out that the tax cuts contributed to an increase in the federal deficit and, therefore, are likely to have distributional effects over many years and across several generations. Whether these changes increased or decreased fairness in the society is ultimately a normative question.

Methodology

The method by which the Census Bureau computes income shares has been challenged by some observers. For example, quintiles of households do not contain the same number of people. Rea Hederman of the Heritage Foundation, a conservative think tank, notes that the top 20% of households contains about 25% of the population. Starting in 2006, the Census Bureau report began calculating a measure called "equivalence-adjusted income" to take into account family size. The Gini coefficient for 2010 using this adjustment fell slightly from 0.469 to 0.457. The trend over time in the two Gini coefficients is similar. Two other flaws pointed out by Mr. Hederman are that taxes and benefits from noncash programs that help the poor are not included. While some Census studies attempt to take these into account and report lower inequality, other studies do not receive as much attention as the main Census annual report.

Even studies that look at incomes over a decade may not capture lifetime income. For example, people in retirement may have a low income but their consumption may be bolstered by drawing on their savings. Younger people may be borrowing to go to school, buy a house, or for other things. The annual income approach of the Census data does not capture this and even the ten-year look in the mobility study mentioned above is too short a period.

This suggests that more precise measurements may provide more insight into explaining inequality.

KEY TAKEAWAYS

- The distribution of income can be illustrated with a Lorenz curve. If all households had the same income, the Lorenz curve would be a 45° line. In general, the more equal the distribution of income, the closer the Lorenz curve will be to the 45° line. A more bowed out curves shows a less equal distribution. The Gini coefficient is another method for describing the distribution of income.
- The distribution of income has, according to the Census Bureau, become somewhat more unequal in the United States during the past 40 years.
- The degree of mobility up and down the distribution of income appears to have declined in recent years.
- Among the factors explaining increased inequality have been changes in family structure and changes in the demand for labor that have rewarded those with college degrees and have penalized unskilled workers.

Case in Point: Attitudes and Inequality

In a fascinating examination of attitudes in the United States and in continental Western Europe, economists Alberto Alesina of Harvard University and George-Marios Angeletos of the Massachusetts Institute of Technology suggest that attitudes about the nature of income earning can lead to quite different economic systems and outcomes concerning the distribution of income.

The economists cite survey evidence from the World Values Survey, which concludes that 71% of Americans, and only 40% of Europeans, agree with the proposition: "*The poor could become rich if they worked hard enough.*" Further, Americans are much more likely to attribute material success to hard work, while Europeans tend to attribute success to factors such as luck, connections, and even corruption. The result, according to Professors Alesina and Angeletos, is that Americans select a government that is smaller and engages in less redistributive activity than is selected by Europeans. Government in continental Western Europe is 50% larger than in the United States, the tax system in Europe is much more progressive than in the United States, regulation of labor and product markets is more extensive in Europe, and redistributive programs are more extensive in Europe than in the United States. As a result, the income distribution in Europe is much more equal than in the United States.

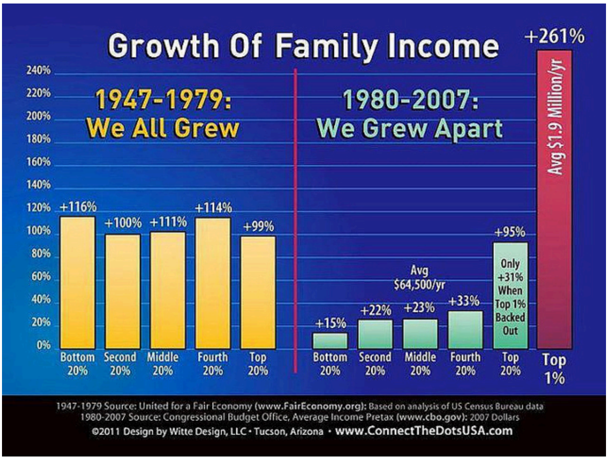
People get what they expect. The economists derive two sets of equilibria. Equilibrium in a society in which people think incomes are a result of luck, connections, and corruption turns out to be precisely that. And, in a society in which people believe incomes are chiefly the result of effort and skill, they are. In the latter society, people work harder and invest more. In the United States, the average worker works 1,600 hours per year. In Europe, the average worker works 1,200 hours per year.

So, who is right—Americans with their "you get what you deserve" or Europeans with their "you get what luck, connections, and corruption bring you" attitude? The two economists show that people get, in effect, what they expect. European values and beliefs produce societies that are more egalitarian. American values and beliefs produce the American result: a society in which the distribution of income is more unequal, the government smaller, and redistribution relatively minor. Professors Alesina and Angeletos conclude that Europeans tend to underestimate the degree to which people can improve their material well-being through hard work, while Americans tend to overestimate that same phenomenon.

320. Reading: Measuring Income Inequality

HOW DO YOU SEPARATE POVERTY AND INCOME INEQUALITY?

Poverty levels can be subjective based on the overall income levels of a country; typically poverty is measured based on a percentage of the median income. Income inequality, however, has to do with the distribution of that income, in terms of which group receives the most or the least income. Income inequality involves comparing those with high incomes, middle incomes, and low incomes—not just looking at those below or near the poverty line. In turn, measuring income inequality means dividing up the population into various groups and then comparing the groups, a task that can be carried out in several ways.



[*Growth in U.S. Family Income: 1947-1979 Versus 1980-2007*] *by Joe Wolf,* [CC-BY-ND].

Poverty can change even when inequality does not move at all. Imagine a situation in which income for everyone in the population declines by 10%. Poverty would rise, since a greater share of the population would now fall below the poverty line. However, inequality would be the same, because everyone suffered the same proportional loss. Conversely, a general rise in income levels over time would keep inequality the same, but reduce poverty.

It is also possible for income inequality to change without affecting the poverty rate. Imagine a situation in which a large number of people who already have high incomes increase their incomes by even more. Inequality would rise as a result—but the number of people below the poverty line would remain unchanged.

Why did inequality of household income increase in the United States in recent decades? Indeed, a trend toward greater income inequality has occurred in many countries around the world, although the effect has been more powerful in the U.S. economy. Economists have focused their explanations for the increasing inequality on two factors that changed more or less continually from the 1970s into the 2000s. One set of explanations focuses on the changing shape of American households; the other focuses on greater inequality of wages, what some economists call "winner take all" *labor markets*. We will begin with how we measure inequality, and then consider the explanations for growing inequality in the United States.

Measuring Income Distribution by Quintiles

One common way of measuring income inequality is to rank all households by income, from lowest to highest, and then to divide all households into five groups with equal numbers of people, known as ***quintiles***. This calculation allows for measuring the distribution of income among the five groups compared to the total. The first quintile is the lowest fifth or 20%, the second quintile is the next lowest, and so on. Income inequality can be measured by comparing what share of the total income is earned by each quintile.

U.S. income distribution by quintile appears in Table 14.5. In 2011, for example, the bottom quintile of the income distribution received 3.2% of income; the second quintile received 8.4%; the third quintile, 14.3%; the fourth quintile, 23.0%; and the top quintile, 51.14%. The final column of Table 14.5 shows what share of income went to households in the top 5% of the income distribution: 22.3% in 2011. Over time, from the late 1960s to the early 1980s, the top fifth of the income distribution typically received between about 43% to 44% of all income. The share of income that the top fifth received then begins to rise. According to the Census Bureau, much of this increase in the share of income going to the top fifth can be traced to an increase in the share of income going to the top 5%. The quintile measure shows how income inequality has increased in recent decades.

Table 14.5. Share of Aggregate Income Received by Each Fifth and Top 5% of Households, 1967–2011 (Source: U.S. Census Bureau, Table H-2)

Year	Lowest Quintile	Second Quintile	Third Quintile	Fourth Quintile	Highest Quintile	Top 5%
1967	4.0	10.8	17.3	24.2	43.6	17.2
1970	4.1	10.8	17.4	24.5	43.3	16.6
1975	4.3	10.4	17.0	24.7	43.6	16.5
1980	4.2	10.2	16.8	24.7	44.1	16.5
1985	3.9	9.8	16.2	24.4	45.6	17.6
1990	3.8	9.6	15.9	24.0	46.6	18.5
1995	3.7	9.1	15.2	23.3	48.7	21.0
2000	3.6	8.9	14.8	23.0	49.8	22.1
2005	3.4	8.6	14.6	23.0	50.4	22.2
2010	3.3	8.5	14.6	23.4	50.3	21.3
2011	3.2	8.4	14.3	23.0	51.1	22.3

It can also be useful to divide the income distribution in ways other than quintiles; for example, into tenths or even into percentiles (that is, hundredths). A more detailed breakdown can provide additional insights. For example, the last column of Table 14.5 shows the income received by the top 5% percent of the income distribution. Between 1980 and 2011, the share of income going to the top 5% increased by 5.8 percentage points (from 16.5% in 1980 to 22.3% in 2011). From 1980 to 2011 the share of income going to the top quintile increased by 7.0 percentage points (from 44.1% in 1980 to 51.1% in 2011). Thus, the top 20% of householders (the fifth quintile) received over half (51.1%) of all the income in the United States in 2011.

Measuring Income Inequality

The U.S. economy has a relatively high degree of income inequality by global standards. As Table 14.7 shows, based on a variety of national surveys done for a selection of years in the last five years of the 2000s (with the exception of Germany, and adjusted to make the measures more comparable), the U.S. economy has greater inequality than Germany (along with most Western European countries). The region of the world with the highest level of income inequality is Latin America, illustrated in the numbers for Brazil and Mexico. The level of inequality in the United States is lower than in some of the low-income countries of the world, like China and Nigeria, or some middle-income countries like the Russian Federation. However, not all poor countries have highly unequal income distributions; India provides a counterexample.

Table 14.7 Income Distribution in Select Countries (Source: U.S. data from U.S. Census Bureau Table H-2. Other data from The World Bank Poverty and Inequality Data Base, http://databank.worldbank.org/data/views/reports/tableview.aspx#)

Country	Survey Year	First Quintile	Second Quintile	Third Quintile	Fourth Quintile	Fifth Quintile
United States	2011	3.2%	8.4%	14.3%	23.0%	51.1%
Germany	2000	8.5%	13.7%	17.8%	23.1%	36.9%
Brazil	2009	2.9%	7.1%	12.4%	19.0%	58.6%
Mexico	2010	4.9%	8.8%	13.3%	20.2%	52.8%
China	2009	4.7%	9.7%	15.3%	23.2%	47.1%
India	2010	8.5%	12.1%	15.7%	20.8%	42.8%
Russia	2009	6.1%	10.4%	14.8%	21.3%	47.1%
Nigeria	2010	4.4%	8.3%	13.0%	20.3%	54.0%

Link It Up

Visit this website to watch a video of wealth inequality across the world.

Lorenz Curve

The data on income inequality can be presented in various ways. For example, you could draw a bar graph that showed the share of income going to each fifth of the income distribution. Figure 14.8 presents an alternative way of showing inequality data in what is called a ***Lorenz curve***. The Lorenz curve shows the cumulative share of population on the horizontal axis and the cumulative percentage of total income received on the vertical axis.

Figure 14.8. *The Lorenz Curve. A Lorenz curve graphs the cumulative shares of income received by everyone up to a certain quintile. The income distribution in 1980 was closer to the perfect equality line than the income distribution in 2011—that is, the U.S. income distribution became more unequal over time.*

Every Lorenz curve diagram begins with a line sloping up at a 45-degree angle, shown as a dashed line in Figure 14.8. The points along this line show what perfect equality of the income distribution looks like. It would mean, for example, that the bottom 20% of the income distribution receives 20% of the total income, the bottom 40% gets 40% of total income, and so on. The other lines reflect actual U.S. data on inequality for 1980 and 2011.

The trick in graphing a Lorenz curve is that you must change the shares of income for each specific quintile, which are shown in the first column of numbers in Table 14.6, into cumulative income, shown in the second column of numbers. For example, the bottom 40% of the cumulative income distribution will be the sum of the first and second quintiles; the bottom 60% of the cumulative income distribution will be the sum of the first, second, and third quintiles, and so on. The final entry in the cumulative income column needs to be 100%, because by definition, 100% of the population receives 100% of the income.

Table 14.6. Calculating the Lorenz Curve

Income Category	Share of Income in 1980 (%)	Cumulative Share of Income in 1980 (%)	Share of Income in 2011 (%)	Cumulative Share of Income in 2011 (%)
First quintile	4.2	4.2	3.2	3.2
Second quintile	10.2	14.4	8.4	11.6
Third quintile	16.8	31.2	14.3	25.9
Fourth quintile	24.7	55.9	23.0	48.9
Fifth quintile	44.1	100.0	51.1	100.0

In a Lorenz curve diagram, a more unequal distribution of income will loop farther down and away from the 45-degree line, while a more equal distribution of income will move the line closer to the 45-degree line. The greater inequality of the U.S. income distribution between 1980 and 2011 is illustrated in Figure 14.8 because the Lorenz curve for 2011 is farther from the 45-degree line than the Lorenz curve for 1980. The Lorenz curve is a useful way of presenting the quintile data that provides an image of all the quintile data at once. The next section shows how income inequality differs in various countries compared to the United States.

321. Reading: Government Policies to Reduce Income Inequality

Government Policies to Reduce Income Inequality

No society should expect or desire complete equality of income at a given point in time, for a number of reasons. First, most workers receive relatively low earnings in their first few jobs, higher earnings as they reach middle age, and then lower earnings after retirement. Thus, a society with people of varying ages will have a certain amount of income inequality. Second, people's preferences and desires differ. Some are willing to work long hours to have income for large houses, fast cars and computers, luxury vacations, and the ability to support children and grandchildren.

These factors all imply that a snapshot of inequality in a given year does not provide an accurate picture of how people's incomes rise and fall over time. Even if some degree of economic inequality is expected at any point in time, how much inequality should there be? There is also the difference between income and wealth, as shown in the following paragraphs.

HOW DO YOU MEASURE

WEALTH VERSUS INCOME INEQUALITY?

Income is a flow of money received, often measured on a monthly or an annual basis; wealth is the sum of the value of all assets, including money in bank accounts, financial investments, a pension fund, and the value of a home. In calculating wealth all debts must be subtracted, such as debt owed on a home mortgage and on credit cards. A retired person, for example, may have relatively little income in a given year, other than a pension or Social Security. However, if that person has saved and invested over time, the person's accumulated wealth can be quite substantial.

The wealth distribution is more unequal than the income distribution, because differences in income can accumulate over time to make even larger differences in wealth. However, the degree of inequality in the wealth distribution can be measured with the same tools we use to measure the inequality in the income distribution, like quintile measurements. Data on wealth are collected once every three years in the Survey of Consumer Finance.

Even if they cannot answer the question of how much inequality is too much, economists can still play an important role in spelling out policy options and tradeoffs. If a society decides to reduce the level of economic inequality, it has three main sets of tools: redistribution from those with high incomes to those with low incomes; trying to assure that a ladder of opportunity is widely available; and a tax on inheritance.

Redistribution

Redistribution means taking income from those with higher incomes and providing income to those with lower incomes. Earlier in this module, we considered some of the key government policies that provide support for the poor: the welfare program TANF, the earned income tax credit, SNAP, and Medicaid. If a reduction in inequality is desired, these programs could receive additional funding.

The programs are paid for through the federal income tax, which is a *progressive tax system* designed in such a way that the rich pay a higher percent in income taxes than the poor. Data from household income tax returns in 2009 shows that the top 1% of households had an average income of $1,219,700 per year in pre-tax income and paid an average federal tax rate of 28.9%. The *effective income tax*, which is total taxes paid divided by total income (all sources of income such as wages, profits, interest, rental income, and government transfers such as veterans' benefits), was much lower. The effective tax paid by the top 1% of householders was 20.4%, while the bottom two quintiles actually paid negative effective income taxes, because of provisions like the earned income tax credit. News stories occasionally report on a high-income person who has managed to pay very little in taxes, but while such individual cases exist, according to the Congressional Budget Office, the typical pattern is that people with higher incomes pay a higher average share of their income in federal income taxes.

Of course, the fact that some degree of redistribution occurs now through the federal income tax and government antipoverty programs does not settle the questions of how much redistribution is appropriate, and whether more redistribution should occur.

The Ladder of Opportunity

Economic inequality is perhaps most troubling when it is not the result of effort or talent, but instead is determined by the circumstances under which a child grows up. One child attends a well-run grade school and high school and heads on to college, while parents help out by supporting education and other interests, paying for college, a first car, and a first house, and offering work connections that lead to internships and jobs. Another child attends a poorly run grade school, barely makes it through a low-quality high school, does not go to college, and lacks family and peer support. These two children may be similar in their underlying talents and in the effort they put forth, but their economic outcomes are likely to be quite different.

Public policy can attempt to build a ladder of opportunities so that, even though all children will never come from identical families and attend identical schools, each child has a reasonable opportunity to attain an economic niche in society based on their interests, desires, talents, and efforts. Some of those initiatives include those shown in Table 14.8.

Table 14.8. Public Policy Initiatives

Children	College Level	Adults
• Improved day care	• Widespread loans and grants for those in financial need	• Opportunities for retraining and acquiring new skills
• Enrichment programs for preschoolers	• Public support for a range of institutions from two-year community colleges to large research universities	• Prohibiting discrimination in job markets and housing on the basis of race, gender, age, and disability
• Improved public schools	–	–
• After school and community activities	–	–
• Internships and apprenticeships	–	–

The United States has often been called a land of opportunity. Although the general idea of a ladder of opportunity for all citizens continues to exert a powerful attraction, specifics are often quite controversial. Society can experiment with a wide variety of proposals for building a ladder of opportunity, especially for those who otherwise seem likely to start their lives in a disadvantaged position. Such policy experiments need to be carried out in a spirit of open-mindedness, because some will succeed while others will not show positive results or will cost too much to enact on a widespread basis.

Inheritance Taxes

There is always a debate about inheritance taxes. It goes like this: On the one hand, why should people who have worked hard all their lives and saved up a substantial nest egg not be able to give their money and possessions to their children and grandchildren?

In particular, it would seem un-American if children were unable to inherit a family business or a family home. On the other hand, many Americans are far more comfortable with inequality resulting from high-income people who earned their money by starting innovative new companies than they are with inequality resulting from high-income people who have inherited money from rich parents.

The United States does have an *estate tax*—that is, a tax imposed on the value of an inheritance—which suggests a willingness to limit how much wealth can be passed on as an inheritance. However, according to the Center on Budget and Policy Priorities, in 2013 the estate tax applied only to those leaving inheritances of more than $5.25 million and thus applies to only a tiny percentage of those with high levels of wealth.

The Tradeoff between Incentives and Income Equality

Government policies to reduce poverty or to encourage economic equality, if carried to extremes, can injure incentives for economic output. The poverty trap, for example, defines a situation where guaranteeing a certain level of income can eliminate or reduce the incentive to work. An extremely high degree of redistribution, with very high taxes on the rich, would be likely to discourage work and entrepreneurship. Thus, it is common to draw the tradeoff between economic output and equality, as shown in Figure 14.10 (a). In this formulation, if society wishes a high level of economic output, like point A, it must also accept a high degree of inequality. Conversely, if society wants a high level of equality, like point B, it must accept a lower level of economic output because of reduced incentives for production.

This view of the tradeoff between economic output and equality may be too pessimistic, and Figure 14.10 (b) presents an alternate vision. Here, the tradeoff between economic output and equality first slopes up, in the vicinity of choice C, suggesting that certain programs might increase both output and economic equality. For example, the policy of providing free public education has an element of redistribution, since the value of the public schooling received by children of low-income families is clearly higher than what low-income families pay in taxes. A well-educated population, however, is also an enormously powerful factor in providing the skilled workers of tomorrow and helping the economy to grow and expand. In this case, equality and economic growth may complement each other.

Moreover, policies to diminish inequality and soften the hardship of poverty may sustain political support for a market economy. After all, if society does not make some effort toward reducing inequality and poverty, the alternative might be that people would rebel against market forces. Citizens might seek economic security by demanding that their legislators pass laws forbidding employers from ever laying off workers or reducing wages, or laws that would impose price floors and price ceilings and shut off international trade. From this viewpoint, policies to reduce inequality may help economic output by building social support for allowing markets to operate.

The graph on the left shows an inverted downward slope with points A and B. The graph on the right shows a more severe inverted downward slope with points C, D, E, F.

Figure 14.10. *The Tradeoff between Incentives and Economic Equality. (a) Society faces a trade-off where any attempt to move toward greater equality, like moving from choice A to B, involves a reduction in economic output. (b) Situations can arise like point C, where it is possible both to increase equality and also to increase economic output, to a choice like D. It may also be possible to increase equality with little impact on economic output, like the movement from choice D to E. However, at some point, too aggressive a push for equality will tend to reduce*

economic output, as in the shift from E to F.

The tradeoff in Figure 14.10 (b) then flattens out in the area between points D and E, which reflects the pattern that a number of countries that provide similar levels of income to their citizens—the United States, Canada, the nations of the European Union, Japan, Australia—have different levels of inequality. The pattern suggests that countries in this range could choose a greater or a lesser degree of inequality without much impact on economic output. Only if these countries push for a much higher level of equality, like at point F, will they experience the diminished incentives that lead to lower levels of economic output. In this view, while a danger always exists that an agenda to reduce poverty or inequality can be poorly designed or pushed too far, it is also possible to discover and design policies that improve equality and do not injure incentives for economic output by very much—or even improve such incentives.

LINK IT UP

OCCUPY WALL STREET

The Occupy movement took on a life of its own over the last few months of 2011, bringing to light issues faced by many people on the lower end of the income distribution. The contents of this module indicate that there is a significant amount of income inequality in the United States. The question is: What should be done about it?

The Great Recession of 2008–2009 caused unemployment to rise and incomes to fall. Many people attribute the recession to mismanagement of the financial system by bankers and financial managers—those in the 1% of the income distribution—but those in lower quintiles bore the greater burden of the recession through unemployment. This seemed to present the picture of inequality in a different light: the group that seemed responsible for the recession was not the group that seemed to bear the burden of the decline in output. A burden shared can bring a society closer together; a burden pushed off onto others can polarize it.

On one level, the problem with trying to reduce income inequality comes down to whether you still believe in the American Dream. If you believe that one day you will have your American Dream—a large income, large house, happy family, or whatever else you would like to have in life—then you do not necessarily want to prevent anyone else from living out their dream. You certainly would not want to run the risk that someone would want to take part of your dream away from you. So there is some reluctance to engage in a redistributive policy to reduce inequality.

However, when those for whom the likelihood of living the American Dream is very small are considered, there are sound arguments in favor of trying to create greater balance. As the text indicated, a little more income equality, gained through long-term programs like increased education and job training, can increase overall economic output. Then everyone is made better off. And the 1% will not seem like such a small group any more.

Self Check: Measuring and Understanding the Distribution of Income

Answer the question(s) below to see how well you understand the

topics covered in the previous section. This short quiz does **not** count toward your grade in the class, and you can retake it an unlimited number of times.

You'll have more success on the Self Check if you've completed the six Readings in this section.

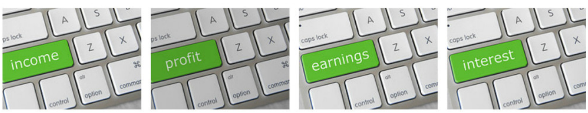
Use this quiz to check your understanding and decide whether to (1) study the previous section further or (2) move on to the next section.

https://assessments.lumenlearning.com/assessments/701

322. Putting It Together: Income Distribution

Summary

The goal of this module was to teach you to understand and evaluate how markets for factors of production affect society's distribution of income.

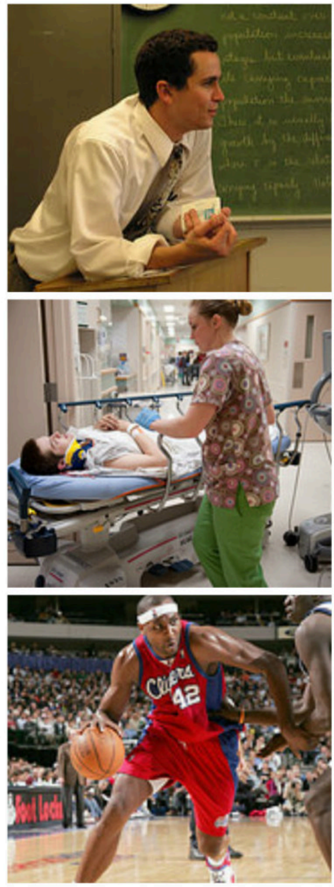


Income, Earnings, Profit, Interest *by* GotCredit, CC-BY.

You learned how to:

- Describe the incomes earned by the factors of production (land, labor, capital, entrepreneurship), i.e. wages, interest, rents, and profit
- Analyze how perfect/imperfect competition between buyers and sellers of factors can impact wages, interest, and rents
- Compare the marginal productivity theory of income distribution versus real world income distribution
- Use the Lorenz Curve to analyze the distribution of income and wealth

Examples



Teacher by Sleepinyourhat, CC-BY-NC-SA. Faces of HCMC by Daniel Guy, CC-BY-NC-ND. Elton Brand by moneyjk89man, CC-BY.

Let's return to some of the questions raised in the Why it Matters feature.

Teachers and nurses are paid less than professional athletes because the market values the former less than the latter. In other words, our actions say that we are willing to pay professional athletes more than teachers and nurses. This may be because athletes are employed through the private sector while most teachers and nurses are employed by the public sector where the lack of market forces makes it harder for workers to be paid what they're worth. Either way, it's a statement about social values.

Urban sanitation engineers get paid a decent wage, not because of the skills required for the job, but rather because of the difficult working conditions in summer and winter. Less "desirable" jobs have to pay more to attract workers.

Unionized workers earn more than non-union workers because unions are able to take advantage of monopoly power in the labor market. Just as a monopoly in the output market can charge a higher price than would be charged if the market were competitive, so unions can charge a higher wage.

The following video explains part of the reason for the increase in income inequality in the U.S in recent years.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://library.achievingthedream.org/sacmicroeconomics/?p=355#oembed-1

323. Glossary: Income Distribution

earned income tax credit (EITC)
: a method of assisting the working poor through the tax system

effective income tax
: percentage of total taxes paid divided by total income

estate tax
: a tax imposed on the value of an inheritance

income inequality
: when one group receives a disproportionate share of total income or wealth than others

income
: a flow of money received, often measured on a monthly or an annual basis

Lorenz curve
: a graph that compares the cumulative income actually received to a perfectly equal distribution of income; it shows the share of population on the horizontal axis and the cumulative percentage of total income received on the vertical axis

Medicaid
: a federal–state joint program enacted in 1965 that provides medical insurance for certain (not all) low-income people, including the near-poor as well as those below the poverty line, and focusing on low-income families with children, the low-income elderly, and the disabled

near-poor
: those who have incomes just above the poverty line

poverty line

the specific amount of income needed for a basic standard of living

poverty rate

percentage of the population living below the poverty line

poverty trap

antipoverty programs set up so that government benefits decline substantially as people earn more income—as a result, working provides little financial gain

poverty

the situation of being below a certain level of income needed for a basic standard of living

progressive tax system

a tax system in which the rich pay a higher percentage of their income in taxes, rather than a higher absolute amount

quintile

dividing a group into fifths, a method often used to look at distribution of income

redistribution

taking income from those with higher incomes and providing income to those with lower incomes

Supplemental Nutrition Assistance Program (SNAP)

a federally funded program, started in 1964, in which each month poor people receive SNAP cards they can use to buy food

safety net

the group of government programs that provide assistance to the poor and the near-poor

wealth

the sum of the value of all assets, including money in bank accounts, financial investments, a pension fund, and the value of a home

324. Discussion: Income Distribution

How would you define income inequality? How is income inequality measured? What has happened to income inequality in the United States since the end of the Second World War in 1945? What evidence can you provide to support your answer? What are the reasons for the changes in income inequality since 1945.

Suggestion: Click here to visit the Web site for the documentary film *Inequality for All*. The site enables you to rent or purchase the film, if you choose. Or you can click on "Graphics Package" to download a free copy of informative graphics from the film. Take a look at those. Summarize what you learn, and share your thoughts.