INTRODUCTION TO ECONOMIC ANALYSIS



Anonymous
LibreTexts



Introduction to Economic Analysis

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 06/13/2026

TABLE OF CONTENTS

Licensing

1: What Is Economics?

- 1.1: Normative and Positive Theories
- 1.2: Opportunity Cost
- 1.3: Economic Reasoning and Analysis

2: Supply and Demand

- 2.1: Demand and Consumer Surplus
- 2.2: Supply and Profit
- 2.3: Market Demand and Supply
- 2.4: Equilibrium
- 2.5: Changes in Demand and Supply

3: Quantification

- 3.1: Elasticity
- 3.2: Supply and Demand Changes

4: The U.S. Economy

- 4.1: Basic Demographics
- 4.2: Education
- 4.3: Households and Consumption
- 4.4: Production
- 4.5: Government
- 4.6: Trade
- 4.7: Fluctuations

5: Government Interventions

- 5.1: Effects of Taxes
- 5.2: Incidence of Taxes
- 5.3: Excess Burden of Taxation
- 5.4: Price Floors and Ceilings
- 5.5: The Politics of Price Controls
- 5.6: Price Supports
- 5.7: Quantity Restrictions and Quotas

6: Trade

- 6.1: Production Possibilities Frontier
- 6.2: Comparative and Absolute Advantage
- 6.3: Factors of Production
- 6.4: International Trade

7: Externalities

- 7.1: External Effects
- 7.2: Pigouvian Taxes
- 7.3: Quotas
- 7.4: Tradable Permits and Auctions
- 7.5: Coasian Bargaining
- 7.6: Fishing and Extinction

8: Public Goods

- 8.1: Free Riders
- 8.2: Provision With Taxation
- 8.3: Local Public Goods

9: Producer Theory- Costs

- 9.1: Types of Firms
- 9.2: Production Functions
- 9.3: Profit Maximization
- 9.4: The Shadow Value
- 9.5: Input Demand
- 9.6: Myriad Costs

10: Producer Theory- Dynamics

- 10.1: Reactions of Competitive Firms
- 10.2: Economies of Scale and Scope
- 10.3: Dynamics With Constant Costs
- 10.4: General Long-run Dynamics

11: Investment

- 11.1: Present Value
- 11.2: Investment
- 11.3: Investment Under Uncertainty
- 11.4: Resource Extraction
- 11.5: A Time to Harvest
- 11.6: Collectibles
- 11.7: Summer Wheat

12: Consumer Theory

- 12.1: Utility Maximization
- 12.2: Budget or Feasible Set
- 12.3: Isoquants
- 12.4: Examples
- 12.5: Substitution Effects
- 12.6: Income Effects
- 12.7: Mathematical Cleanup

13: Applied Consumer Theory

- 13.1: Labor Supply
- 13.2: Urban Real Estate Prices

- 13.3: Dynamic Choice
- 13.4: Risk Aversion
- 13.5: Search

14: General Equilibrium

- 14.1: Edgeworth Box
- 14.2: Equilibrium With Price System
- 14.3: General Equilibrium

15: Monopoly

- 15.1: Sources of Monopoly
- 15.2: Basic Analysis
- 15.3: Effect of Taxes
- 15.4: Price Discrimination
- 15.5: Welfare Effects
- 15.6: Natural Monopoly
- 15.7: Peak-load Pricing

16: Games and Strategic Behavior

- 16.1: Matrix Games
- 16.2: Nash Equilibrium
- 16.3: Mixed Strategies
- 16.4: Examples
- 16.5: Subgame Perfection
- 16.6: Supergames

17: Imperfect Competition

- 17.1: Cournot Oligopoly
- 17.2: Cournot Industry Performance
- 17.3: Hotelling Differentiation
- 17.4: The Circle Model

18: Information

- 18.1: Market for Lemons
- 18.2: Myerson-Satterthwaite Theorem
- 18.3: Signaling
- 18.4: Search and Price Dispersion

19: Agency Theory

- 19.1: Principals and Agents
- 19.2: Cost of Providing Incentives
- 19.3: Multi-tasking without Homogeneity
- 19.4: Multi-tasking without Homogeneity

20: Auctions

- 20.1: English Auction
- 20.2: Sealed-bid Auction
- 20.3: Dutch Auction
- 20.4: Vickrey Auction

- 20.5: The Winner’s Curse and Linkage
- 20.6: Auction Design

21: Antitrust

- 21.1: Sherman Act
- 21.2: Clayton Act
- 21.3: Price Fixing
- 21.4: Mergers

Index

Index

Glossary

Detailed Licensing

Detailed Licensing

Licensing LicensingEdit Update titlecancel

This page is a draft and is under active development.

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: What Is Economics?

1.1: Normative and Positive Theories

1.2: Opportunity Cost

1.3: Economic Reasoning and Analysis

This page titled 1: What Is Economics? is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

1.1: Normative and Positive Theories 1.1: Normative and Positive TheoriesEdit Update titlecancel

Learning ObjectiveS

1. How is economics used?
2. What is an economic theory?
3. What is a market?

Economic analysis serves two main purposes. The first is to understand how goods and services, the scarce resources of the economy, are actually allocated in practice. This is a **positive analysis**, like the study of electromagnetism or molecular biology; it aims to understand the world without value judgments. The development of this positive theory, however, suggests other uses for economics. Economic analysis can predict how changes in laws, rules, and other government policies will affect people and whether these changes are socially beneficial on balance. Such predictions combine positive analysis—predicting the effects of changes in rules—with studies that make value judgments known as **normative analyses**. For example, a gasoline tax to build highways harms gasoline buyers (who pay higher prices) but helps drivers (by improving the transportation system). Since drivers and gasoline buyers are typically the same people, a normative analysis suggests that everyone will benefit. Policies that benefit everyone are relatively uncontroversial.

In contrast, **cost-benefit analysis** weighs the gains and losses to different individuals to determine changes that provide greater benefits than harm. For example, a property tax to build a local park creates a benefit to park users but harms property owners who pay the tax. Not everyone benefits, since some taxpayers don't use the park. Cost-benefit analysis weighs the costs against the benefits to determine if the policy is beneficial on balance. In the case of the park, the costs are readily measured in monetary terms by the size of the tax. In contrast, the benefits are more difficult to estimate. Conceptually, the benefits are the amount the park users would be willing to pay to use the park. However, if there is no admission charge to the park, one must estimate a **willingness-to-pay**, the amount a customer is willing and able to pay for a good. In principle, the park provides greater benefits than costs if the benefits to the users exceed the losses to the taxpayers. However, the park also involves transfers from one group to another.

Welfare analysis is another approach to evaluating government intervention into markets. It is a normative analysis that trades off gains and losses to different individuals. Welfare analysis posits social preferences and goals, such as helping the poor. Generally a welfare analysis requires one to perform a cost-benefit analysis, which accounts for the overall gains and losses but also weighs those gains and losses by their effects on other social goals. For example, a property tax to subsidize the opera might provide more value than costs, but the bulk of property taxes are paid by lower- and middle-income people, while the majority of operagoers are wealthy. Thus, the opera subsidy represents a transfer from relatively low-income people to wealthy people, which contradicts societal goals of equalization. In contrast, elimination of sales taxes on basic food items like milk and bread has a greater benefit to the poor, who spend a much larger percentage of their income on food, than do the rich. Thus, such schemes are desirable primarily for their redistribution effects. Economics is helpful for providing methods to determining the overall effects of taxes and programs, as well as the distributive impacts. What economics can't do, however, is advocate who ought to benefit. That is a matter for society to decide.

KEY TAKEAWAYS

- A positive analysis, analogous to the study of electromagnetism or molecular biology, involves only the attempt to understand the world around us without value judgments.
- Economic analyses employing value judgments are known as normative analyses. When everyone is made better off by a change, recommending that change is relatively uncontroversial.
- A cost-benefit analysis totals the gains and losses to different individuals in dollars and suggests carrying out changes that provide greater benefits than harm. A cost-benefit analysis is a normative analysis.
- Welfare analysis posits social preferences and goals, permitting an optimization approach to social choice. Welfare analysis is normative.
- Economics helps inform society about the consequences of decisions, but the valuation of those decisions is a matter for society to choose.

This page titled 1.1: Normative and Positive Theories is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

1.2: Opportunity Cost 1.2: Opportunity CostEdit Update titlecancel

Learning ObjectiveS

1. What is opportunity cost?
2. How is it computed?
3. What is its relationship to the usual meaning of cost?

Economists think of cost in a slightly quirky way that makes sense, however, once you think about it for a while. We use the term **opportunity cost** to remind you occasionally of our idiosyncratic notion of cost. For an economist, the cost of buying or doing something is the value that one forgoes in purchasing the product or undertaking the activity of the thing. For example, the cost of a university education includes the tuition and textbook purchases, as well as the wages that were lost during the time the student was in school. Indeed, the value of the time spent in acquiring the education is a significant cost of acquiring the university degree. However, some "costs" are not opportunity costs. Room and board would not be a cost since one must eat and live whether one is working or at school. Room and board are a cost of an education only insofar as they are expenses that are only incurred in the process of being a student. Similarly, the expenditures on activities that are precluded by being a student—such as hang-gliding lessons, or a trip to Europe—represent savings. However, the value of these activities has been lost while you are busy reading this book.

Opportunity cost is defined by the following:

The opportunity cost is the value of the best forgone alternative.

This definition emphasizes that the cost of an action includes the monetary cost as well as the value forgone by taking the action. The opportunity cost of spending $19 to download songs from an online music provider is measured by the benefit that you would have received had you used the $19 instead for another purpose. The opportunity cost of a puppy includes not just the purchase price but the food, veterinary bills, carpet cleaning, and time value of training as well. Owning a puppy is a good illustration of opportunity cost, because the purchase price is typically a negligible portion of the total cost of ownership. Yet people acquire puppies all the time, in spite of their high cost of ownership. Why? The economic view of the world is that people acquire puppies because the value they expect exceeds their opportunity cost. That is, they reveal their preference for owning the puppy, as the benefit they derive must apparently exceed the opportunity cost of acquiring it.

Even though opportunity costs include nonmonetary costs, we will often monetize opportunity costs, by translating these costs into dollar terms for comparison purposes. Monetizing opportunity costs is valuable, because it provides a means of comparison. What is the opportunity cost of 30 days in jail? It used to be that judges occasionally sentenced convicted defendants to "thirty days or thirty dollars," letting the defendant choose the sentence. Conceptually, we can use the same idea to find out the value of 30 days in jail. Suppose you would pay a fine of $750 to avoid the 30 days in jail but would serve the time instead to avoid a fine of $1,000. Then the value of the 30-day sentence is somewhere between $750 and $1,000. In principle there exists a critical price at which you're indifferent to "doing the time" or "paying the fine." That price is the monetized or dollar cost of the jail sentence.

The same process of selecting between payment and action may be employed to monetize opportunity costs in other contexts. For example, a gamble has a **certainty equivalent**, which is the amount of money that makes one indifferent to choosing the gamble versus the certain payment. Indeed, companies buy and sell risk, and the field of **risk management** is devoted to studying the buying or selling of assets and options to reduce overall risk. In the process, risk is valued, and the riskier stocks and assets must sell for a lower price (or, equivalently, earn a higher average return). This differential, known as a **risk premium**, is the monetization of the risk portion of a gamble.

Buyers shopping for housing are presented with a variety of options, such as one- or two-story homes, brick or wood exteriors, composition or shingle roofing, wood or carpet floors, and many more alternatives. The approach economists adopt for valuing these items is known as **hedonic pricing**. Under this method, each item is first evaluated separately and then the item values are added together to arrive at a total value for the house. The same approach is used to value used cars, making adjustments to a base value for the presence of options like leather interior, GPS system, iPod dock, and so on. Again, such a valuation approach converts a bundle of disparate attributes into a monetary value.

The conversion of costs into dollars is occasionally controversial, and nowhere is it more so than in valuing human life. How much is your life worth? Can it be converted into dollars? Some insight into this question can be gleaned by thinking about risks. Wearing seatbelts and buying optional safety equipment reduce the risk of death by a small but measurable amount. Suppose a $400 airbag reduces the overall risk of death by 0.01%. If you are indifferent to buying the airbag, you have implicitly valued the

probability of death at $400 per 0.01%, or $40,000 per 1%, or around $4,000,000 per life. Of course, you may feel quite differently about a 0.01% chance of death compared with a risk 10,000 times greater, which would be a certainty. But such an approach provides one means of estimating the value of the risk of death—an examination of what people will, and will not, pay to reduce that risk.

KEY TAKEAWAYS

- The opportunity cost is the value of the best-forgone alternative.
- Opportunity cost of a purchase includes more than the purchase price but all of the costs associated with a choice.
- The conversion of costs into dollar terms, while sometimes controversial, provides a convenient means of comparing costs.

This page titled 1.2: Opportunity Cost is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

1.3: Economic Reasoning and Analysis 1.3: Economic Reasoning and AnalysisEdit Update titlecancel

Learning ObjectiveS

1. How do economists reason?
2. What is comparative static?
3. What assumptions are commonly made by economists about human behavior?
4. What do economists mean by marginal?

What this country needs is some one-armed economists.

—Harry S. Truman

Economic reasoning is rather easy to satirize. One might want to know, for instance, what the effect of a policy change—a government program to educate unemployed workers, an increase in military spending, or an enhanced environmental regulation—will be on people and their ability to purchase the goods and services they desire. Unfortunately, a single change may have multiple effects. As an absurd and tortured example, government production of helium for (allegedly) military purposes reduces the cost of children's birthday balloons, causing substitution away from party hats and hired clowns. The reduction in demand for clowns reduces clowns' wages and thus reduces the costs of running a circus. This cost reduction increases the number of circuses, thereby forcing zoos to lower admission fees to compete with circuses. Thus, were the government to stop subsidizing the manufacture of helium, the admission fees of zoos would likely rise, even though zoos use no helium. This example is superficially reasonable, although the effects are miniscule.

To make any sense of all the effects of a change in economic conditions, it is helpful to divide up the effects into pieces. Thus, we will often look at the effects of a change in relation to "other things equal," that is, assuming nothing else has changed. This isolates the effect of the change. In some cases, however, a single change can lead to multiple effects; even so, we will still focus on each effect individually. A gobbledygook way of saying "other things equal" is to use Latin and say "**ceteris paribus**." Part of your job as a student is to learn economic jargon, and that is an example. Fortunately, there isn't too much jargon.

We will make a number of assumptions that you may find implausible. Not all of the assumptions we make are necessary for the analysis, but instead are used to simplify things. Some, however, are necessary and therefore deserve an explanation. There is a frequent assumption in economics that the people we will talk about are exceedingly selfish relative to most people we know. We model the choices that people make, presuming that they select on the basis of their own welfare only. Such people—the people in the models as opposed to real people—are known as "**homo economicus**." Real people are indubitably more altruistic than homo economicus, because they couldn't be less: homo economicus is entirely selfish. (The technical term is **self-interested behavior**.) That doesn't necessarily invalidate the conclusions drawn from the theory, however, for at least four reasons:

1. People often make decisions as families or households rather than as individuals, and it may be sensible to consider the household as the "consumer." Identifying households as fairly selfish is more plausible perhaps than identifying individuals as selfish.
2. Economics is mostly silent on why consumers want things. You may wish to make a lot of money to build a hospital or endow a library, which would be altruistic. Such motives are not inconsistent with self-interested behavior.
3. Corporations are expected to serve their shareholders by maximizing share value, thus inducing self-interested behavior on the part of the corporation. Even if corporations could ignore the interests of their shareholders, capital markets would require them to consider shareholder interests as necessary condition for raising funds to operate and invest. In other words, people choosing investments for high returns will force corporations to seek a high return.
4. There are good, as well as bad, consequences that follow from people acting in their self-interest, and it is important for us to know what they are.

Thus, while the theory of self-interested behavior may not be universally descriptive, it is nonetheless a good starting point for building a framework to study the economics of human behavior.

Self-interested behavior will often be described as "maximizing behavior," where consumers maximize the value they obtain from their purchases, and firms maximize their profits. One objection to this economic methodology is that people rarely carry out the calculations necessary to literally maximize anything. However, that is not a fatal flaw to the methodology. People don't consciously do the physics calculations to throw a baseball or thread a needle, yet they somehow accomplish these tasks. Economists often consider that people act "as if" they maximize an objective, even though no explicit calculation is performed.

Some corporations in fact use elaborate computer programs to minimize costs or maximize profits, and the field of operations research creates and implements such maximization programs. Thus, while individuals don't necessarily calculate the consequences of their behavior, some companies do.

A good example of economic reasoning is the **sunk cost fallacy**. Once one has made a significant nonrecoverable investment, there is a psychological tendency to invest more, even when subsequent investment isn't warranted. France and Britain continued to invest in the Concorde (a supersonic aircraft no longer in production) long after they realized that the project would generate little return. If you watch a movie to the end, even after you know it stinks, you haven fallen prey to the sunk cost fallacy. The fallacy is attempting to make an investment that has gone bad turn out to be good, even when it probably won't. The popular phrase associated with the sunk cost fallacy is "throwing good money after bad." The fallacy of sunk costs arises because of a psychological tendency to make an investment pay off when something happens to render it obsolete. It is a mistake in many circumstances.

Casinos often exploit the fallacy of sunk costs. People who lose money gambling hope to recover their losses by gambling more. The sunk "investment" to win money may cause gamblers to invest even more in order to win back what has already been lost. For most games like craps, blackjack, and one-armed bandits, the house wins on average, so that the average gambler (and even the most skilled slot machine or craps player) loses on average. Thus, for most, trying to win back losses is to lose more on average.

The way economics performs is by a proliferation of mathematical models, and this proliferation is reflected in this book. Economists reason with models. Models help by removing extraneous details from a problem or issue, which allows one more readily to analyze what remains. In some cases the models are relatively simple, like supply and demand. In other cases, the models are more complex. In all cases, the models are constructed to provide the simplest analysis possible that allows us to understand the issue at hand. The purpose of the model is to illuminate connections between ideas. A typical implication of a model is "when *A* increases, *B* falls." This "**comparative static**" prediction lets us determine how *A* affects *B*, at least in the setting described by the model. The real world is typically much more complex than the models we postulate. That doesn't invalidate the model, but rather by stripping away extraneous details, the model is a lens for focusing our attention on specific aspects of the real world that we wish to understand.

One last introductory warning before we get started. A parody of economists talking is to add the word **marginal** before every word. Marginal is just economists' jargon for "the derivative of." For example, marginal cost is the derivative of cost; marginal value is the derivative of value. Because introductory economics is usually taught to students who have not yet studied calculus (or can't be trusted to remember it), economists avoid using derivatives and instead refer to the value of the next unit purchased, or the cost of the next unit, in terms of the marginal value or cost. This book uses "marginal" frequently because we wish to introduce the necessary jargon to students who want to read more advanced texts or take more advanced classes in economics. For an economics student not to know the word marginal would be akin to a physics student who does not know the word mass. The book minimizes jargon where possible, but part of the job of a principled student is to learn the jargon, and there is no getting around that.

KEY TAKEAWAYS

- It is often helpful to break economic effects into pieces.
- A common strategy is to examine the effects of a change in relation to "other things equal," that is, assuming nothing else has changed, which isolates the effect of the change. "Ceteris paribus" means "other things equal."
- Economics frequently models the choices that people make by assuming that they make the best choice for them. People in a model are known occasionally as "homo economicus." Homo economicus is entirely selfish. The technical term is acting in one's self-interest.
- Self-interested behavior is also described as "maximizing behavior," where consumers maximize the net value they obtain from their purchases, and firms maximize their profits.
- Once one has made a significant nonrecoverable investment, there is a psychological tendency to invest more, even when the return on the subsequent investment isn't worthwhile, known as the sunk cost fallacy.
- Economists reason with models. By stripping out extraneous details, the model represents a lens to isolate and understand aspects of the real world.
- Marginal is just economists' jargon for "the derivative of." For example, marginal cost is the derivative of cost; marginal value is the derivative of value.

This page titled 1.3: Economic Reasoning and Analysis is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

2: Supply and Demand

2.1: Demand and Consumer Surplus
2.2: Supply and Profit
2.3: Market Demand and Supply
2.4: Equilibrium
2.5: Changes in Demand and Supply

This page titled 2: Supply and Demand is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

2.1: Demand and Consumer Surplus 2.1: Demand and Consumer SurplusEdit Update titlecancel

Learning Objectives

- What is demand?
- What is the value to buyers of their purchases?
- What assumptions are commonly made about demand?
- What causes demand to rise or fall?
- What is a good you buy only because you are poor?
- What are goods called that are consumed together?
- How does the price of one good influence demand for other goods?

Eating a french fry makes most people a little bit happier, and most people are willing to give up something of value—a small amount of money or a little bit of time—to eat one. The personal value of the french fry is measured by what one is willing to give up to eat it. That value, expressed in dollars, is the willingness to pay for french fries. So, if you are willing to give up 3 cents for a single french fry, your willingness to pay is 3 cents. If you pay a penny for the french fry, you've obtained a net of 2 cents in value. Those 2 cents—the difference between your willingness to pay and the amount you pay—is known as **consumer surplus**. Consumer surplus is the value of consuming a good, minus the price paid.

The value of items—like french fries, eyeglasses, or violins—is not necessarily close to what one must pay for them. For people with bad vision, eyeglasses might be worth $10,000 or more in the sense that people would be willing to pay this amount or more to wear them. Since one doesn't have to pay nearly this much for eyeglasses means that the consumer surplus derived from eyeglasses is enormous. Similarly, an order of french fries might be worth $3 to a consumer, but since they are available for $1, the consumer obtains a surplus of $2 from purchase.

How much is a second order of french fries worth? For most of us, the first order is worth more than the second one. If a second order is worth $2, we would still gain from buying it. Eating a third order of fries is worth less still, and at some point we're unable or unwilling to eat any more fries even when they are free, that implies that the value of additional french fries becomes zero eventually.

We will measure consumption generally as units per period of time, for example, french fries consumed per month.

Many, but not all, goods have this feature of **diminishing marginal value**—the value of the last unit declines as the number consumed rises. If we consume a quantity q, that implies the marginal value, denoted by v (q), falls as the number of units rise.When marginal value falls, which may occur with beer consumption, constructing demand takes some additional effort, which isn't a great deal of consequence. Buyers will still choose to buy a quantity where marginal value is decreasing. An example is illustrated in Figure 2.1, where the value is a straight line, declining in the number of units.

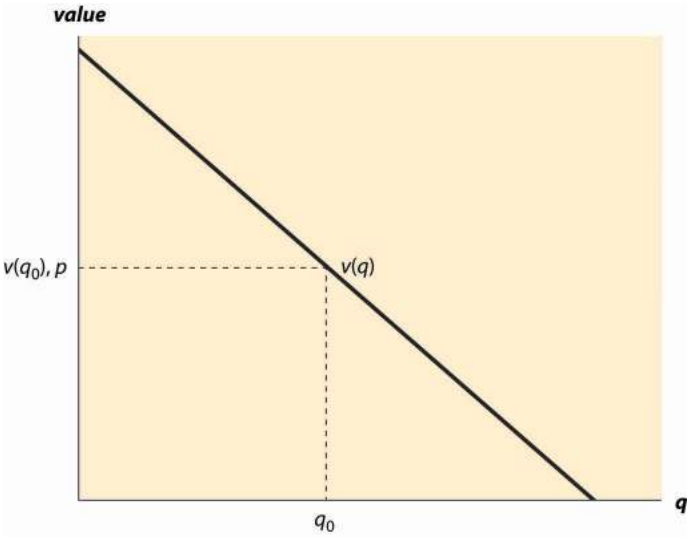


Figure 2.1 The demand curve

Demand needn't be a straight line, and indeed could be any downward-sloping curve. Contrary to the usual convention, the quantity demanded for any price is represented by the vertical axis whereas the price is plotted along the horizontal.

It is often important to distinguish the demand curve—the relationship between price and quantity demanded—from the quantity demanded. Typically, "demand" refers to the curve, while "quantity demanded" is a point on the curve.

For a price p, a consumer will buy units q such that v (q) > p since those units are worth more than they cost. Similarly, a consumer would not buy units for which v (q) < p. Thus, the quantity q_0 that solves the equation v (q_0) = p indicates the quantity the consumer will buy. This value is illustrated in Figure 2.1.We will treat units as continuous, even though they are discrete units. This simplifies the mathematics; with discrete units, the consumer buys those units with value exceeding the price and doesn't buy those with value less than the price, just as before. However, since the value function isn't continuous, much less differentiable, it would be an accident for marginal value to equal price. It isn't particularly difficult to accommodate discrete products, but it doesn't enhance the model so we opt for the more convenient representation. Another way of expressing this insight is that the marginal value curve is the inverse of the demand function, where the demand function gives the quantity purchased at a given price. Formally, if x (p) is the quantity a consumer buys at price p, then v(x(p))=p.

But what is the marginal value curve? Suppose the total value of consumption is u(q). A consumer who pays u(q) for the quantity q is indifferent to receiving nothing and paying nothing. For each quantity, there should exist one and only one price that makes the consumer indifferent between purchasing and receiving nothing. If the consumer is just willing to pay u (q), any additional amount exceeds what the consumer should be willing to pay.

The consumer facing price p receives consumer surplus of $CS = u(q) - pq$. In order to obtain the maximal benefit, the consumer chooses q to maximize u (q) – pq. When the function CS is maximized, its derivative is zero. This implies that the quantity maximizing the consumer surplus must satisfy

$$0 = ddq(u(q) - pq) = u'(q) - p$$

Thus, $v(q) = u'(q)$; implying that the marginal value is the derivative of the total value.

Consumer surplus is the value of the consumption minus the amount paid, and it represents the net value of the purchase to the consumer. Formally, it is u (q) – pq. A graph of consumer surplus is generated by the following identity:

$$CS = maxq(u(q) - pq) = u(q0) - pq0 = \int 0q0(u'(x) - p)dx = \int 0q0(v(x) - p)dx.$$

This expression shows that consumer surplus can be represented as the area below the demand curve and above the price, as illustrated in Figure 2.2. The consumer surplus represents the consumer's gains from trade, the value of consumption to the consumer net of the price paid.

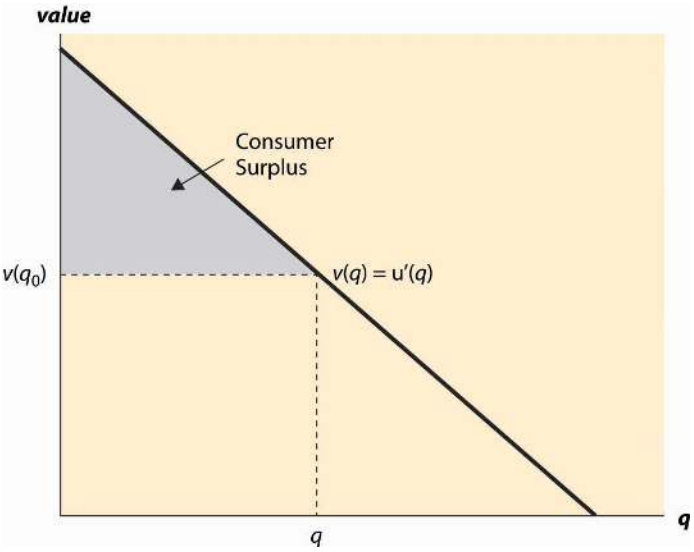


Figure 2.2 Consumer surplus

The consumer surplus can also be expressed using the demand curve, by integrating from the price up to where the demand curve intersects with the price axis. In this case, if x(p) is demand, we have

$CS = \int p \infty x(y)\, dy$.

When you buy your first car, you experience an increase in demand for gasoline because gasoline is pretty useful for cars and not so much for other things. An imminent hurricane increases the demand for plywood (to protect windows), batteries, candles, and bottled water. An increase in demand is represented by a movement of the entire curve to the northeast (up and to the right), which represents an increase in the marginal value v (movement up) for any given unit, or an increase in the number of units demanded for any given price (movement to the right). Figure 2.3 illustrates a shift in demand.

Similarly, the reverse movement represents a decrease in demand. The beauty of the connection between demand and marginal value is that an increase in demand could, in principle, have meant either more units demanded at a given price or a higher willingness to pay for each unit, but those are in fact the same concept. Both changes create a movement up and to the right.

For many goods, an increase in income increases the demand for the good. Porsche automobiles, yachts, and Beverly Hills homes are mostly purchased by people with high incomes. Few billionaires ride the bus. Economists aptly named goods whose demand doesn't increase with income **inferior goods**, with the idea that people substitute to better quality, more expensive goods as their incomes rise. When demand for a good increases with income, the good is called a **normal good**. It would have been better to call such goods superior, but it is too late to change such a widely accepted convention.

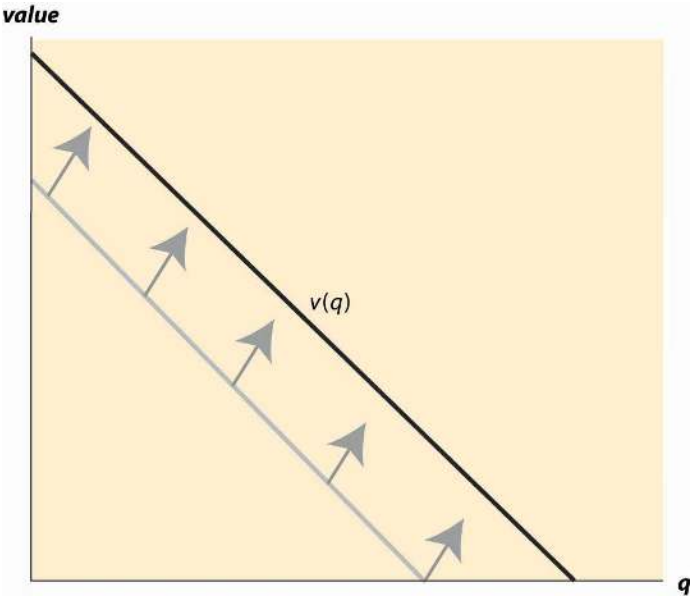


Figure 2.3 An increase in demand

Another factor that influences demand is the price of related goods. The dramatic fall in the price of computers over the past 20 years has significantly increased the demand for printers, monitors, and Internet access. Such goods are examples of **complements**. Formally, for a given good x, a complement is a good whose consumption increases the value of x. Thus, the use of computers increases the value of peripheral devices like printers and monitors. The consumption of coffee increases the demand for cream for many people. Spaghetti and tomato sauce, national parks and hiking boots, air travel and hotel rooms, tables and chairs, movies and popcorn, bathing suits and sunscreen, candy and dentistry—all are examples of complements for most people. Consumption of one increases the value of the other. The complementary relationship is typically symmetric—if consumption of x increases the value of y, then consumption of y must increase the value of x.The basis for this insight can be seen by denoting the total value in dollars of consuming goods x and y as u(x, y). Then the demand for x is given by the partial ∂u ∂x . The statement that y is a complement means that the demand for x rises as y increases; that is, $\partial 2u \partial x \partial y > 0$. But then with a continuous second derivative, $\partial 2u \partial y \partial x > 0$, which means the demand for y, ∂u ∂y , increases with x. From this we can predict that if the price of good y decreases, then the amount good y, a complementary good to x, will decline. Why, you may ask? The reason is that consumers will purchase more of good x when its price decreases. This will make good y more valuable, and hence consumers will also purchase more of good y as a result.

The opposite case of a complement is a **substitute**. For a given good x, a substitute is a good whose consumption decreases the value of x. Colas and root beer are substitutes, and a fall in the price of root beer (resulting in an increase in the consumption of root beer) will tend to decrease the demand for colas. Pasta and ramen, computers and typewriters, movies (in theaters) and sporting events, restaurants and dining at home, spring break in Florida versus spring break in Mexico, marijuana and beer, economics courses and psychology courses, driving and bicycling—these are all examples of substitutes for most people. An increase in the price of a substitute increases the demand for a good; and, conversely, a decrease in the price of a substitute decreases demand for a good. Thus, increased enforcement of the drug laws, which tends to increase the price of marijuana, leads to an increase in the demand for beer.

Much of demand is merely idiosyncratic to the individual—some people like plaids, some like solid colors. People like what they like. People often are influenced by others—tattoos are increasingly common, not because the price has fallen but because of an increased acceptance of body art. Popular clothing styles change, not because of income and prices but for other reasons. While there has been a modest attempt to link clothing style popularity to economic factors,Skirts are allegedly shorter during economic booms and lengthen during recessions. by and large there is no coherent theory determining fads and fashions beyond the observation that change is inevitable. As a result, this course, and economics generally, will accept preferences for what they are without questioning why people like what they like. While it may be interesting to understand the increasing social acceptance of tattoos, it is beyond the scope of this text and indeed beyond most, but not all, economic analyses. We will, however, account for some of the effects of the increasing acceptance of tattoos through changes in the number of parlors offering tattooing, changes in the variety of products offered, and so on.

Key Takeaways

- Demand is the function that gives the number of units purchased as a function of the price.
- The difference between your willingness to pay and the amount you pay is known as consumer surplus. Consumer surplus is the value in dollars of a good minus the price paid.
- Many, but not all, goods have the feature of diminishing marginal value—the value of the last unit consumed declines as the number consumed rises.
- Demand is usually graphed with price on the vertical axis and quantity on the horizontal axis.
- Demand refers to the entire curve, while quantity demanded is a point on the curve.
- The marginal value curve is the inverse of demand function.
- Consumer surplus is represented in a demand graph by the area between demand and price.
- An increase in demand is represented by a movement of the entire curve to the northeast (up and to the right), which represents an increase in the marginal value v (movement up) for any given unit, or an increase in the number of units demanded for any given price (movement to the right). Similarly, the reverse movement represents a decrease in demand.
- Goods whose demand doesn't increase with income are called inferior goods, with the idea that people substitute to better quality, more expensive goods as their incomes rise. When demand for a good increases with income, the good is called normal.
- Demand is affected by the price of related goods.
- For a given good x, a complement is a good whose consumption increases the value of x. The complementarity relationship is symmetric—if consumption of x increases the value of y, then consumption of y must increase the value of x.
- The opposite case of a complement is a substitute. An increase in the consumption of a substitute decreases the value for a good.

EXERCISES

1. A reservation price is is a consumer's maximum willingness to pay for a good that is usually bought one at a time, like cars or computers. Graph the demand curve for a consumer with a reservation price of $30 for a unit of a good.
2. Suppose the demand curve is given by $x(p) = 1 - p$. The consumer's expenditure is $p * x(p) = p(1 - p)$. Graph the expenditure. What price maximizes the consumer's expenditure?
3. For demand $x(p) = 1 - p$, compute the consumer surplus function as a function of p.
4. For demand $x(p) = p^{-}\varepsilon$, for ε > 1, find the consumer surplus as a function of p. (Hint: Recall that the consumer surplus can be expressed as $CS = \int p \infty x(y)dy$.)
5. Suppose the demand for wheat is given by $qd = 3–p$ and the supply of wheat is given by $qs = 2p$, where p is the price.
    1. Solve for the equilibrium price and quantity.
    2. Graph the supply and demand curves. What are the consumer surplus and producer profits?
    3. Now suppose supply shifts to $qs = 2p + 1$. What are the new equilibrium price and quantity?
6. How will the following affect the price of a regular cup of coffee, and why?
    1. Droughts in Colombia and Costa Rica
    2. A shift toward longer work days
    3. The price of milk falls
    4. A new study that shows many great health benefits of tea
7. A reservation price is a consumer's maximum willingness to pay for a good that is usually bought one at a time, like cars or computers. Suppose in a market of T-shirts, 10 people have a reservation price of $10 and the 11th person has a reservation price of $5. What does the demand "curve" look like?
8. In Exercise 7, what is the equilibrium price if there were 9 T-shirts available? What if there were 11 T-shirts available? How about 10?
9. A consumer's value for slices of pizza is given by the following table. Graph this person's demand for slices of pizza.

Slices of pizza	Total value
0	0
1	4
2	7

Slices of pizza	Total value
3	10
4	12
5	11

This page titled 2.1: Demand and Consumer Surplus is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

2.2: Supply and Profit 2.2: Supply and ProfitEdit Update titlecancel

Learning Objectives

- What is supply?
- What are gains made by sellers called?
- What assumptions are commonly made about supply?
- What causes supply to rise or fall?
- What are goods produced together called?
- How do the prices of one good influence supply for other goods?

The term **supply** refers to the function that gives the quantity offered for sale as a function of price. The supply curve gives the number of units that will be supplied on the horizontal axis, as a function of the price on the vertical axis; Figure 2.4 illustrates a supply curve. Generally, supply is upward sloping, because if it is a good deal for a supplier to sell 50 units of a product at a price of $10, then it is an even better deal to supply those same 50 at a price of $11. The seller might choose to sell more than 50, but if the first 50 aren't worth keeping at a price of $10, then it remains true at $11.This is a good point at which to remind the reader that the economists' familiar assumption of "other things equal" is still in effect. If the increased price is an indication that prices might rise still further, or a consequence of some other change that affects the seller's value of items, then of course the higher price might not justify sale of the items. We hold other things equal to focus on the effects of price alone, and then will consider other changes separately. The pure effect of an increased price should be to increase the quantity offered, while the effect of increased expectations may be to decrease the quantity offered.

The seller with cost c (q) of selling q units obtains a profit, at price p per unit, of $pq - c(q)$. The quantity that maximizes profit for the seller is the quantity q* satisfying $0 = ddqpq - c(q) = p - c'(q^*)$

Thus, "price equals marginal cost" is a characteristic of profit maximization; the supplier sells all the units whose cost is less than price, and doesn't sell the units whose cost exceeds price. In constructing the demand curve, we saw that it was the inverse of the marginal value. There is an analogous property of supply: The supply curve is the inverse function of marginal cost. Graphed with the quantity supplied on the horizontal axis and price on the vertical axis, the supply curve is the marginal cost curve, with marginal cost on the vertical axis.

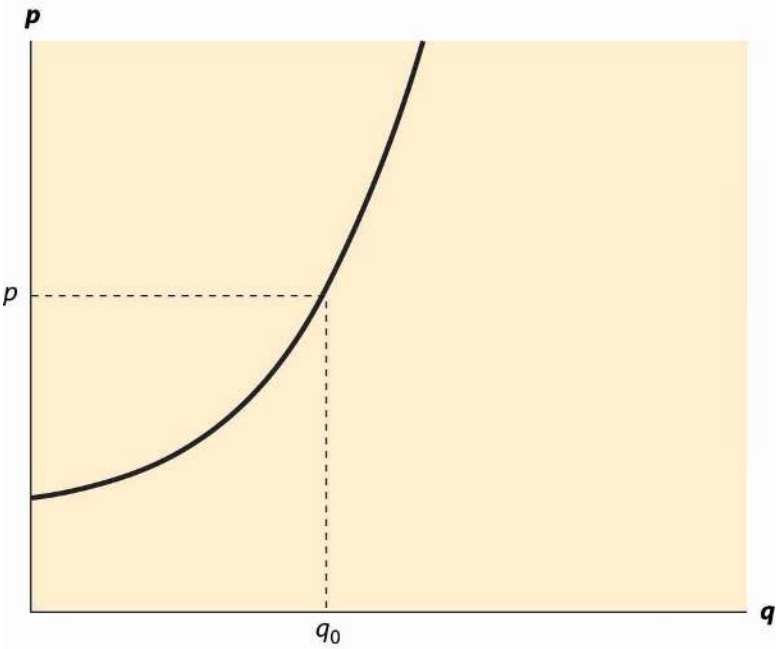


Figure 2.4 The supply curve

Analogous to consumer surplus with demand, profit is given by the difference of the price and marginal cost.

$$\text{Profit } = \max qpq - c(q) = pq^* - c(q^*) = \int 0q^* (p - c'(x))\, dx$$

This area is shaded in Figure 2.5.

The relationship of demand and marginal value exactly parallels the relationship of supply and marginal cost, for a somewhat hidden reason. Supply is just negative demand; that is, a supplier is just the possessor of a good who doesn't keep it but instead, offers it to the market for sale. For example, when the price of housing goes up, one of the ways people demand less is by offering to rent a room in their house—that is, by supplying some of their housing to the market. Similarly, the marginal cost of supplying a good already produced is the loss of not having the good—that is, the marginal value of the good. Thus, with exchange, it is possible to provide the theory of supply and demand entirely as a theory of net demand, where sellers are negative demanders. There is some mathematical economy in this approach, and it fits certain circumstances better than separating supply and demand. For example, when the price of electricity rose very high in the western United States in 2003, several aluminum smelters resold electricity that they had purchased in long-term contracts; in other words, demanders became suppliers.

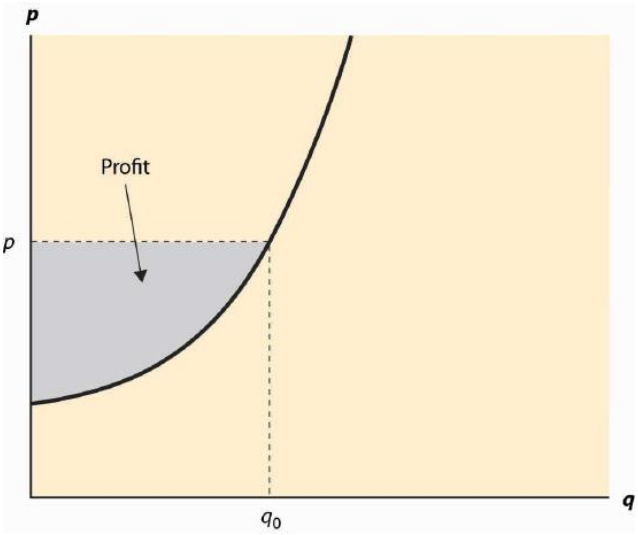


Figure 2.5 Supplier profits

However, the "net demand" approach obscures the likely outcomes in instances where the sellers are mostly distinct from the buyers. Moreover, while there is a theory of complements and substitutes for supply that is exactly parallel to the equivalent theory for demand, the nature of these complements and substitutes tends to be different. For these reasons, and also for the purpose of being consistent with common economic usage, we will distinguish supply and demand.

Figure 2.6 An increase in supply

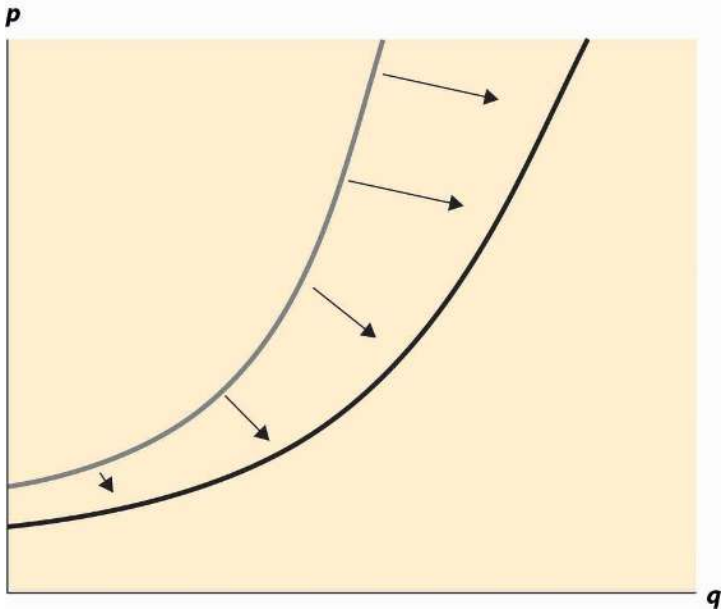


Figure 2.6. A decrease in supply is the reverse case, a shift to the northwest.

Anything that increases costs of production will tend to increase marginal cost and thus reduce the supply. For example, as wages rise, the supply of goods and services is reduced because wages are the input price of labor. Labor accounts for about two thirds of all input costs, and thus wage increases create supply reductions (a higher price is necessary to provide the same quantity) for most goods and services. Costs of materials, of course, increase the price of goods using those materials. For example, the most important input into the manufacture of gasoline is crude oil, and an increase of $1 in the price of a 42-gallon barrel of oil increases the price of gasoline about 2 cents—almost one-for-one by volume. Another significant input in many industries is capital and, as we will see, interest is the cost of capital. Thus, increases in interest rates increase the cost of production, and thus tend to decrease the supply of goods.

Analogous to complements in demand, a complement in supply to a good x is a good y such that an increase in the production of y increases the supply of x. In demand, a **complement in supply** is a good whose cost falls as the amount produced of another good rises. Complements in supply are usually goods that are jointly produced. In producing lumber (sawn boards), a large quantity of wood chips and sawdust are also produced as a by-product. These wood chips and sawdust are useful in the manufacture of paper. An increase in the price of lumber tends to increase the quantity of trees sawn into boards, thereby increasing the supply of wood chips. Thus, lumber and wood chips are complements in supply.

It turns out that copper and gold are often found in the same kinds of rock—the conditions that give rise to gold compounds also give rise to copper compounds. Thus, an increase in the price of gold tends to increase the number of people prospecting for gold and, in the process, increases not just the quantity of gold supplied to the market but also the quantity of copper. Thus, copper and gold are complements in supply.

The classic supply–complement is beef and leather—an increase in the price of beef increases the slaughter of cows, thereby increasing the supply of leather.

The opposite of a complement in supply is a **substitute in supply**. This is a good whose cost rises as the amount produced of another good rises. Military and civilian aircraft are substitutes in supply—an increase in the price of military aircraft will tend to divert resources used in the manufacture of aircraft toward military aircraft and away from civilian aircraft, thus reducing the supply of civilian aircraft. Wheat and corn are also substitutes in supply. An increase in the price of wheat will lead farmers whose land is well suited to producing either wheat or corn to substitute wheat for corn, thus increasing the quantity of wheat and decreasing the quantity of corn. Agricultural goods grown on the same type of land are usually substitutes. Similarly, cars and trucks, tables and desks, sweaters and sweatshirts, horror movies and romantic comedies are all examples of substitutes in supply.

Complements and substitutes are important because they are common and have predictable effects on demand and supply. Changes in one market spill over to the other market through the mechanism of complements or substitutes.

Key Takeaways

- The supply curve gives the number of units as a function of the price that will be supplied for sale to the market.
- Price equals marginal cost is an implication of profit maximization; the supplier sells all the units whose cost is less than price and doesn't sell the units whose cost exceeds price.
- The supply curve is the inverse function of marginal cost. Graphed with the quantity supplied on the horizontal axis and price on the vertical axis, the supply curve is the marginal cost curve, with marginal cost on the vertical axis.
- Profit is given by the difference of the price and marginal cost.
- Supply is negative demand.
- An increase in supply refers to either more units available at a given price or a lower price for the supply of the same number of units. Thus, an increase in supply is graphically represented by a curve that is lower or to the right, or both—that is, to the southeast. A decrease in supply is the reverse case, a shift to the northwest.
- Anything that increases costs of production will tend to increase marginal cost and thus reduce the supply.
- A complement in supply to a good x is a good y such that an increase in the price of y increases the supply of x.
- The opposite of a complement in supply is a substitute in supply.

EXERCISES

1. A typist charges $30 per hour and types 15 pages per hour. Graph the supply of typed pages.
2. An owner of an oil well has two technologies for extracting oil. With one technology, the oil can be pumped out and transported for $5,000 per day, and 1,000 barrels per day are produced. With the other technology, which involves injecting natural gas into the well, the owner spends $10,000 per day and $5 per barrel produced, but 2,000 barrels per day are produced. What is the supply? Graph it.

   (Hint: Compute the profits, as a function of the price, for each of the technologies. At what price would the producer switch from one technology to the other? At what price would the producer shut down and spend nothing?)

3. An entrepreneur has a factory that produces L^{α} widgets, where $\alpha < 1$, when L hours of labor are used. The cost of labor (wage and benefits) is w per hour. If the entrepreneur maximizes profit, what is the supply curve for widgets?

   (Hint: The entrepreneur's profit, as a function of the price, is $pL^{a} - wL$. The entrepreneur chooses the amount of labor to maximize profit. Find the amount of labor that maximizes profit, which is a function of p, w, and α. The supply is the amount of output produced, which is L^{α}.)

4. In the above exercise, suppose now that more than 40 hours entails a higher cost of labor (overtime pay). Let w be $20/hr for under 40 hours, and $30/hr for each hour over 40 hours, and $\alpha = ½$. Find the supply curve.

   (Hint: Let L(w, p) be the labor demand when the wage is w (no overtime pay) and the price is p. Now show that, if $L(20, p) < 40$, the entrepreneur uses $L(20, p)$hours. This is shown by verifying that profits are higher at $L(20, p)$ than at $L(30, p)$. If $L(30, p) > 40$, the entrepreneur uses $L(30, p)$hours. Finally, if $L(20, p) > 40 > L(30, p)$, the entrepreneur uses 40 hours. Labor translates into supply via Lα.)

5. In the previous exercise, for what range of prices does employment equal 40 hours? Graph the labor demanded by the entrepreneur.

   (Hint: The answer involves 10 .)

6. Suppose marginal cost, as a function of the quantity q produced, is mq. Find the producer's profit as a function of the price p.

This page titled 2.2: Supply and Profit is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

2.3: Market Demand and Supply 2.3: Market Demand and SupplyEdit Update titlecancel

Learning Objectives

- How are individual demands and supplies aggregated to create a market?

Individuals with their own supply or demand trade in a market, where prices are determined. Markets can be specific or virtual locations—the farmers' market, the New York Stock Exchange, eBay—or may be an informal or more amorphous market, such as the market for restaurant meals in Billings, Montana, or the market for roof repair in Schenectady, New York.

Figure 2.7 Market demand

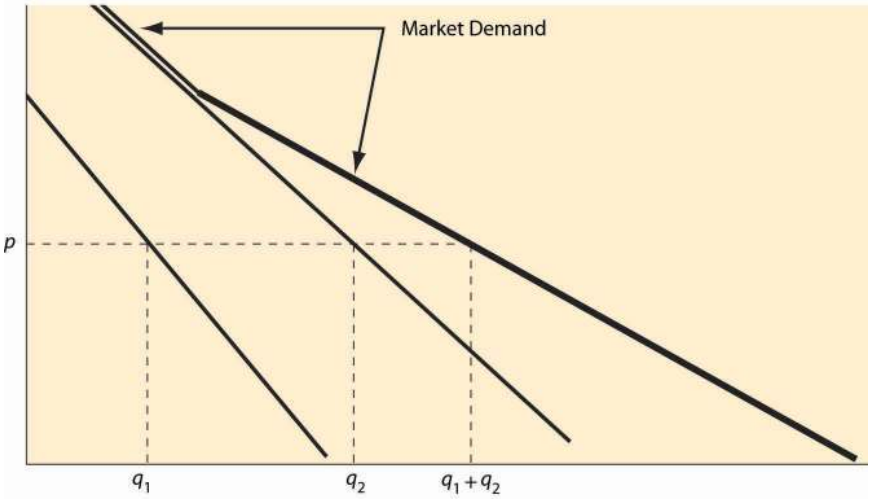


Figure 2.7. For a given price p, the quantity q_1 demanded by one consumer and the quantity q_2 demanded by a second consumer are illustrated. The sum of these quantities represents the market demand if the market has just those two participants. Since the consumer with subscript 2 has a positive quantity demanded for high prices, while the consumer with subscript 1 does not, the market demand coincides with consumer 2's demand when the price is sufficiently high. As the price falls, consumer 1 begins purchasing, and the market quantity demanded is larger than either individual participant's quantity and is the sum of the two quantities.

Example: If the demand of Buyer 1 is given by $q = max0, 10–p$, and the demand of Buyer 2 is given by $q = max0, 20–4p$, what is market demand for the two participants?

Solution: First, note that Buyer 1 buys zero at a price of 10 or higher, while Buyer 2 buys zero at a price of 5 or higher. For a price above 10, market demand is zero. For a price between 5 and 10, market demand is Buyer 1's demand, or $10–p$. Finally, for a price between zero and 5, the market quantity demanded is $10–p + 20–4p = 30–5p$.

Market supply is similarly constructed—the market supply is the horizontal (quantity) sum of all the individual supply curves.

Example: If the supply of Firm 1 is given by $q = 2p$, and the supply of Firm 2 is given by $q = max0, 5p–10$, what is market supply for the two participants?

Solution: First, note that Firm 1 is in the market at any price, but Firm 2 is in the market only if price exceeds 2. Thus, for a price between zero and 2, market supply is Firm 1's supply, or 2p. For $p > 2$, market supply is $5p–10 + 2p = 7p–10$.

Key Takeaways

- The market demand gives the quantity purchased by all the market participants—the sum of the individual demands—for each price. This is sometimes called a "horizontal sum" because the summation is over the quantities for each price.
- The market supply is the horizontal (quantity) sum of all the individual supply curves.

EXERCISES

1. Is the consumer surplus for market demand the sum of the consumer surpluses for the individual demands? Why or why not? Illustrate your conclusion with a figure like Figure 2.7.

2. Suppose the supply of firm i is αi p, when the price is p, where i takes on the values $1, 2, 3, \ldots, n$. What is the market supply of these n firms?
3. Suppose consumers in a small town choose between two restaurants, A and B. Each consumer has a value vA for A's meal and a value vB for B's meal, and each value is a uniform random draw from the [0, 1] interval. Consumers buy whichever product offers the higher consumer surplus. The price of B's meal is 0.2. In the square associated with the possible value types, identify which consumers buy from A. Find the demand, which is the area of the set of consumers who buy from A in the diagram below. [Hint: Consumers have three choices—buy nothing [value 0], buy from A [value vA – pA], and buy from B [value vB – pB = vB – 0.2).] Draw the lines illustrating which choice has the highest value for the consumer.

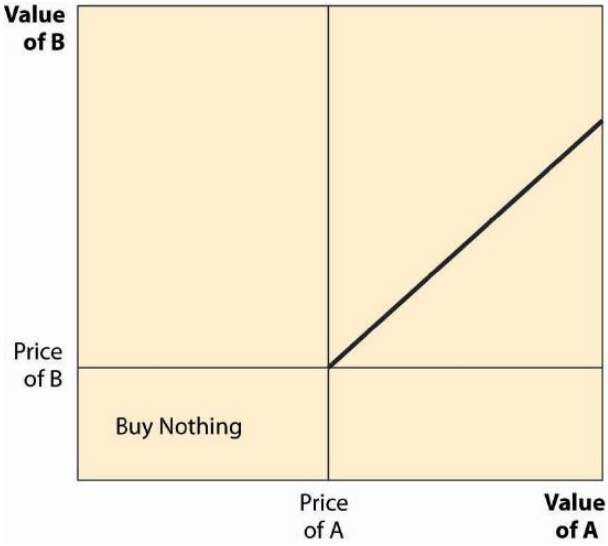


Figure 2.8

This page titled 2.3: Market Demand and Supply is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

2.4: Equilibrium 2.4: EquilibriumEdit Update titlecancel

Learning Objectives

- How are prices determined?
- What happens when price is too low?
- What happens when price is too high?
- When will price remain constant?

Economists use the term equilibrium in the same way that the word is used in physics: to represent a steady state in which opposing forces are balanced so that the current state of the system tends to persist. In the context of supply and demand, **equilibrium** occurs when the pressure for higher prices is balanced by the pressure for lower prices, and so that rate of exchange between buyers and sellers persists.

When the current price is above the equilibrium price, the quantity supplied exceeds the quantity demanded, and some suppliers are unable to sell their goods because fewer units are purchased than are supplied. This condition, where the quantity supplied exceeds the quantity demanded, is called a **surplus**. The suppliers failing to sell have an incentive to offer their good at a slightly lower price—a penny less—to make a sale. Consequently, when there is a surplus, suppliers push prices down to increase sales. In the process, the fall in prices reduces the quantity supplied and increases the quantity demanded, thus eventually eliminating the surplus. That is, a surplus encourages price-cutting, which reduces the surplus, a process that ends only when the quantity supplied equals the quantity demanded.

Similarly, when the current price is lower than the equilibrium price, the quantity demanded exceeds the quantity supplied, and a **shortage** exists. In this case, some buyers fail to purchase, and these buyers have an incentive to offer a slightly higher price to make their desired purchase. Sellers are pleased to receive higher prices, which tends to put upward pressure on the price. The increase in price reduces the quantity demanded and increases the quantity supplied, thereby eliminating the shortage. Again, these adjustments in price persist until the quantity supplied equals the quantity demanded.

Figure 2.9 Equilibration

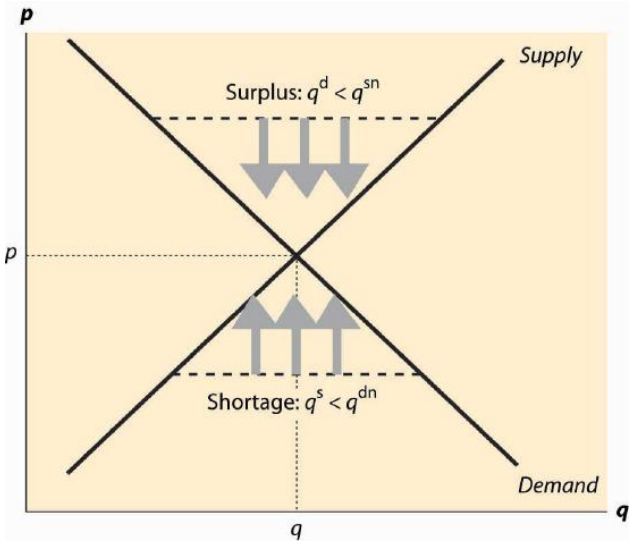


Figure 2.9, a surplus arises at any price above the equilibrium price p*, because the quantity supplied qs is larger than the quantity demanded qd. The effect of the surplus—leading to sellers with excess inventory—induces price-cutting, which is illustrated using three arrows pointing down.

Similarly, when the price is below p*, the quantity supplied qs is less than the quantity demanded qd. This causes some buyers to fail to find goods, leading to higher asking prices and higher bid prices by buyers. The tendency for the price to rise is illustrated using three arrows pointing up. The only price that doesn't lead to price changes is p*, the equilibrium price in which the quantity supplied equals the quantity demanded.

The logic of equilibrium in supply and demand is played out daily in markets all over the world—from stock, bond, and commodity markets with traders yelling to buy or sell, to Barcelona fish markets where an auctioneer helps the market find a price, to Istanbul's gold markets, to Los Angeles's real estate markets.

The equilibrium of supply and demand balances the quantity demanded and the quantity supplied so that there is no excess of either. Would it be desirable, from a social perspective, to force more trade or to restrain trade below this level?

There are circumstances where the equilibrium level of trade has harmful consequences, and such circumstances are considered in the chapter on externalities. However, provided that the only people affected by a transaction are the buyer and the seller, the equilibrium of supply and demand maximizes the total gains from trade.

This proposition is quite easy to see. To maximize the gains from trade, clearly the highest value buyers must get the goods. Otherwise, if a buyer that values the good less gets it over a buyer who values it more, then gains can arise from them trading. Similarly, the lowest-cost sellers must supply those goods; otherwise we can increase the gains from trade by replacing a higher-cost seller with a lower-cost seller. Thus, the only question is how many goods should be traded to maximize the gains from trade, since it will involve the lowest-cost suppliers selling to the highest-value buyers. Adding a trade increases the total gains from trade when that trade involves a buyer with value higher than the seller's cost. Thus, the gains from trade are maximized by the set of transactions to the left of the equilibrium, with the high-value buyers buying from the low-cost sellers.

In the economist's language, the equilibrium is **efficient** in that it maximizes the gains from trade under the assumption that the only people affected by any given transaction are the buyers and the seller.

Key Takeaways

- The quantity supplied of a good or service exceeding the quantity demanded is called a surplus.
- If the quantity demanded exceeds the quantity supplied, a shortage exists.
- The equilibrium price is the price in which the quantity supplied equals the quantity demanded.
- The equilibrium of supply and demand maximizes the total gains from trade.

EXERCISES

1. If demand is given by $q^d \mathrm{a}(p) = a - bp$, and supply is given by $q^s(p) = cp$, solve for the equilibrium price and quantity. Find the consumer surplus and producer profits.
2. If demand is given by $q^d \operatorname{ap} \varepsilon = ap^{-\varepsilon}$, and supply is given by $q^s(p) = bp^n$, where all parameters are positive numbers, solve for the equilibrium price and quantity.

This page titled 2.4: Equilibrium is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

2.5: Changes in Demand and Supply 2.5: Changes in Demand and SupplyEdit Update titlecancel

Learning Objectives

- What are the effects of changes in demand and supply?

What are the effects of an increase in demand? As the population of California has grown, the demand for housing has risen. This has pushed the price of housing up and also spurred additional development, increasing the quantity of housing supplied as well. We see such a demand increase illustrated in Figure 2.10, which represents an increase in the demand. In this figure, supply and demand have been abbreviated *S* and *D*. Demand starts at D_1 and is increased to D_2. Supply remains the same. The equilibrium price increases from p_1* to p_2*, and the quantity rises from q_1* to q_2*.

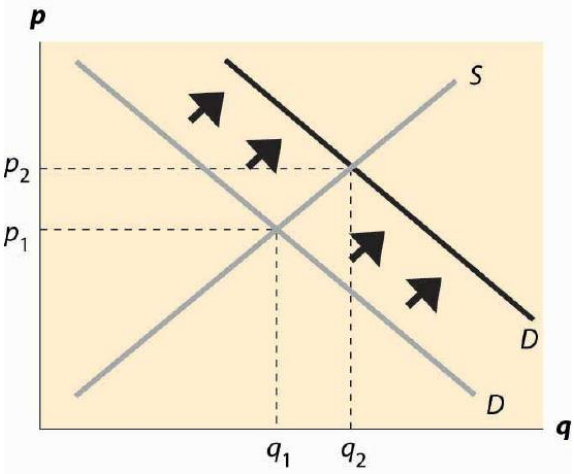


Figure 2.10 An increase in demand

A decrease in demand—which occurred for typewriters with the advent of computers, or buggy whips as cars replaced horses as the major method of transportation—has the reverse effect of an increase and implies a fall in both the price and the quantity traded. Examples of decreases in demand include products replaced by other products—VHS tapes were replaced by DVDs, vinyl records were replaced by CDs, cassette tapes were replaced by CDs, floppy disks (oddly named because the 1.44 MB "floppy," a physically hard product, replaced the 720 KB, 5 ¼-inch soft floppy disk) were replaced by CDs and flash memory drives, and so on. Even personal computers experienced a fall in demand as the market was saturated in 2001.

An increase in supply comes about from a fall in the marginal cost: recall that the supply curve is just the marginal cost of production. Consequently, an increased supply is represented by a curve that is lower and to the right on the supply–demand graph, which is an endless source of confusion for many students. The reasoning—lower costs and greater supply are the same thing—is too easily forgotten. The effects of an increase in supply are illustrated in Figure 2.11. The supply curve goes from S_1 to S_2, which represents a lower marginal cost. In this case, the quantity traded rises from q_1* to q_2* and price falls from p_1* to p_2*.

Figure 2.11 An increase in supply

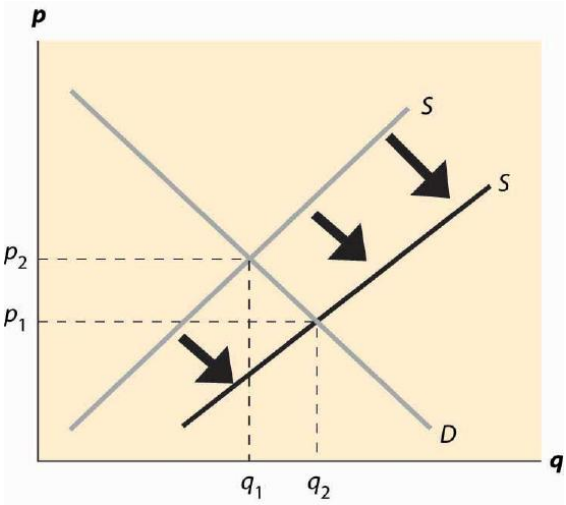


Figure 2.12.Used with permission of computer storage expert, Dr. Edward Grochowski. Note that the prices in this figure reflect a logarithmic scale, so that a fixed-percentage decrease is illustrated by a straight line. Prices of DRAMs fell to close to 1/1,000 of their 1990 level by 2004. The means by which these prices have fallen are themselves quite interesting. The main reasons are shrinking the size of the chip (a "die shrink"), so that more chips fit on each silicon disk, and increasing the size of the disk itself, so again more chips fit on a disk. The combination of these two, each of which required solutions to thousands of engineering and chemistry problems, has led to dramatic reductions in marginal costs and consequent increases in supply. The effect has been that prices fell dramatically and quantities traded rose dramatically.

An important source of supply and demand changes can be found in the markets of complements. A decrease in the price of a demand–complement increases the demand for a product; and, similarly, an increase in the price of a demand–substitute increases the demand for a product. This gives two mechanisms to trace through effects from external markets to a particular market via the linkage of demand substitutes or complements. For example, when the price of gasoline falls, the demand for automobiles (a complement) should increase overall. As the price of automobiles rises, the demand for bicycles (a substitute in some circumstances) should rise. When the price of computers falls, the demand for operating systems (a complement) should rise. This gives an operating system seller like Microsoft an incentive to encourage technical progress in the computer market in order to make the operating system more valuable.

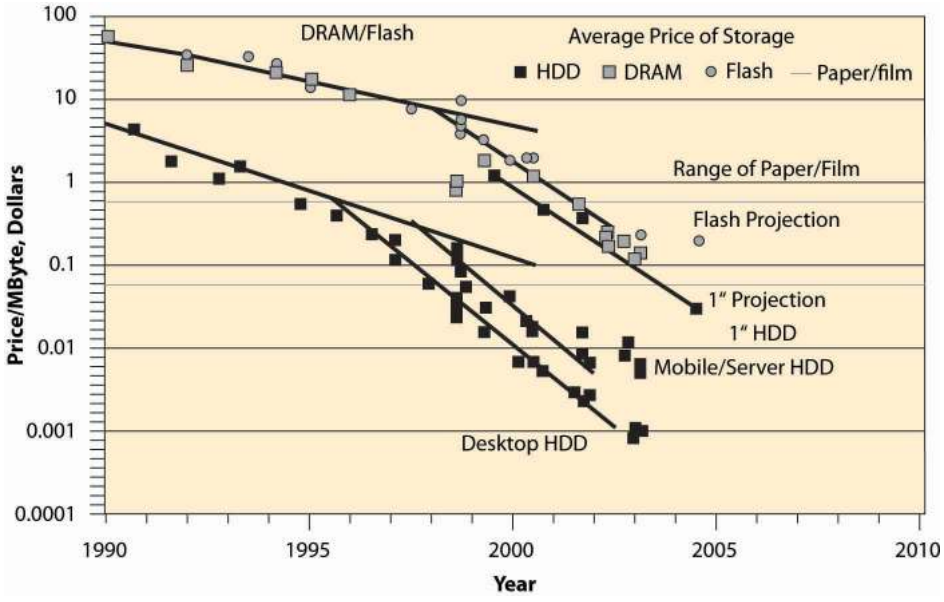


Figure 2.12 Price of storage

An increase in the price of a supply–substitute reduces the supply of a product (by making the alternative good more attractive to suppliers); and, similarly, a decrease in the price of a supply-complement reduces the supply of a good. By making the by-product less valuable, the returns to investing in a good are reduced. Thus, an increase in the price of DVD-R disks (used for recording DVDs) discourages investment in the manufacture of CD-R disks, which are a substitute in supply, leading to a decrease in the

supply of CD-Rs. This tends to increase the price of CD-Rs, other things equal. Similarly, an increase in the price of oil increases exploration for oil, which increases the supply of natural gas, which is a complement in supply. However, since natural gas is also a demand substitute for oil (both are used for heating homes), an increase in the price of oil also tends to increase the demand for natural gas. Thus, an increase in the price of oil increases both the demand and the supply of natural gas. Both changes increase the quantity traded, but the increase in demand tends to increase the price, while the increase in supply tends to decrease the price. Without knowing more, it is impossible to determine whether the net effect is an increase or decrease in the price.

When the price of gasoline goes up, people curtail their driving to some extent but don't immediately scrap their SUVs to buy more fuel-efficient automobiles or electric cars. Similarly, when the price of electricity rises, people don't immediately replace their air conditioners and refrigerators with the most modern, energy-saving ones. There are three significant issues raised by this kind of example. First, such changes may be transitory or permanent, and people react differently to temporary changes than to permanent changes. Second, energy is a modest portion of the cost of owning and operating an automobile or refrigerator, so it doesn't make sense to scrap a large capital investment over a small permanent increase in cost. Thus, people rationally continue to operate "obsolete" devices until their useful life is over, even when they wouldn't buy an exact copy of that device. This situation, in which past choices influence current decisions, is called **hysteresis**. Third, a permanent increase in energy prices leads people to buy more fuel-efficient cars and to replace their old gas-guzzlers more quickly. That is, the effects of a change are larger over a time period long enough that all inputs can be changed (which economists call the **long run**) than over a shorter time interval where not all inputs can be changed, or the **short run**.

A striking example of such delays arose when oil quadrupled in price in 1973 and 1974, caused by a reduction in sales by the cartel of oil-producing nations, *OPEC*, which stands for the Organization of Petroleum Exporting Countries. The increased price of oil (and consequent increase in gasoline prices) caused people to drive less and to lower their thermostats in the winter, thus reducing the quantity of oil demanded. Over time, however, they bought more fuel-efficient cars and insulated their homes more effectively, significantly reducing the quantity demanded still further. At the same time, the increased prices for oil attracted new investments into oil production in Alaska, the North Sea between Britain and Norway, Mexico, and other areas. Both of these effects (long-run substitution away from energy and long-run supply expansion) caused the price to fall over the longer term, undoing the supply reduction created by OPEC. In 1981, OPEC further reduced output, sending prices still higher; but again, additional investments in production, combined with energy-saving investments, reduced prices until they fell back to 1973 levels (adjusted for inflation) in 1986. Prices continued to fall until 1990 (reaching an all-time low level) when Iraq's invasion of Kuwait and the resulting first Iraqi war sent them higher again.

Short-run and long-run effects represent a theme of economics, with the major conclusion of the theme being that substitution doesn't occur instantaneously, which leads to predictable patterns of prices and quantities over time.

It turns out that direct estimates of demand and supply are less useful as quantifications than notions of percentage changes, which have the advantage of being unit-free. This observation gives rise to the concept of elasticity, the next topic.

Key Takeaways

- An increase in the demand increases both the price and quantity traded.
- A decrease in demand implies a fall in both the price and the quantity traded.
- An increase in the supply decreases the price and increases the quantity traded.
- A decrease in the supply increases the price and decreases the quantity traded.
- A change in the supply of a good affects its price. This price change will in turn affect the demand for both demand complements and demand subsitutes.
- People react less to temporary changes than to permanent changes. People rationally continue to operate "obsolete" devices until their useful life is over, even when they wouldn't buy an exact copy of that device, an effect called hysteresis.
- Short-run and long-run effects represent a theme of economics, with the major conclusion that substitution doesn't occur instantaneously, which leads to predictable patterns of prices and quantities over time.

EXERCISES

1. Video games and music CDs are substitutes in demand. What is the effect of an increase in supply of video games on the price and quantity traded of music CDs? Illustrate your answer with diagrams for both markets.
2. Electricity is a major input into the production of aluminum, and aluminum is a substitute in supply for steel. What is the effect of an increase in price of electricity on the steel market?

3. Concerns about terrorism reduced demand for air travel and induced consumers to travel by car more often. What should happen to the price of Hawaiian hotel rooms?

This page titled 2.5: Changes in Demand and Supply is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

3: Quantification

3.1: Elasticity
3.2: Supply and Demand Changes

This page titled 3: Quantification is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

3.1: Elasticity 3.1: ElasticityEdit Update titlecancel

Learning Objectives

- What is the best way of measuring the responsiveness of demand?
- What is the best way of measuring the responsiveness of supply?

Let $x(p)$ represent the quantity purchased when the price is p, so that the function x represents demand. How responsive is demand to price changes? One might be tempted to use the derivative, x', to measure the responsiveness of demand, since it measures how much the quantity demanded changes in response to a small change in price. However, this measure has two problems. First, it is sensitive to a change in units. If I measure the quantity of candy in kilograms rather than in pounds, the derivative of demand for candy with respect to price changes even if the demand itself is unchanged. Second, if I change price units, converting from one currency to another, again the derivative of demand will change. So the derivative is unsatisfactory as a measure of responsiveness because it depends on units of measure. A common way of establishing a unit-free measure is to use percentages, and that suggests considering the responsiveness of demand to a small percentage change in price in percentage terms. This is the notion of **elasticity of demand**. The concept of elasticity was invented by Alfred Marshall (1842–1924) in 1881 while sitting on his roof. The elasticity of demand is the percentage decrease in quantity that results from a small percentage increase in price. Formally, the elasticity of demand, which is generally denoted with the Greek letter epsilon, ε, (chosen mnemonically to indicate elasticity) is

$$\varepsilon = -dxxdpp = -pxdxdp = -px'(p)x(p)$$

The minus sign is included in the expression to make the elasticity a positive number, since demand is decreasing. First, let's verify that the elasticity is, in fact, unit free. A change in the measurement of x doesn't affect elasticity because the proportionality factor appears in both the numerator and denominator. Similarly, a change in the measure of price so that p is replaced by r = ap, does not change the elasticity, since as demonstrated below,

$$\varepsilon = -rddrx(r/a)x(r/a) = -rx'(r/a)1ax(r/a) = -px'(p)x(p)$$

the measure of elasticity is independent of a, and therefore not affected by the change in units.

How does a consumer's expenditure, also known as (individual) total revenue, react to a change in price? The consumer buys x(p) at a price of p, and thus **total expenditure**, or total revenue, is TR = px(p). Thus,

$$d \text{ dp } px(p) = x(p) + px'(p) = x(p)\,(1 + px'(p)x(p)) = x(p)(1 - \varepsilon)$$

Therefore,

$$ddpTR1pTR = 1 - \varepsilon.$$

In other words, the percentage change of total revenue resulting from a 1% change in price is one minus the elasticity of demand. Thus, a 1% increase in price will increase total revenue when the elasticity of demand is less than one, which is defined as an **inelastic demand**. A price increase will decrease total revenue when the elasticity of demand is greater than one, which is defined as an **elastic demand**. The case of elasticity equal to one is called **unitary elasticity**, and total revenue is unchanged by a small price change. Moreover, that percentage increase in price will increase revenue by approximately 1 – ε percent. Because it is often possible to estimate the elasticity of demand, the formulae can be readily used in practice. Table 3.1 provides estimates on demand elasticities for a variety of products.

Table 3.1 Various Demand Elasticities

Product	Product	ε
Salt	Movies	0.9
Matches	Shellfish, consumed at home	0.9
Toothpicks	Tires, short-run	0.9
Airline travel, short-run	Oysters, consumed at home	1.1
Residential natural gas, short-run	Private education	1.1

Product	Product	ε
Gasoline, short-run	Housing, owner occupied, long-run	1.2
Automobiles, long-run	Tires, long-run	1.2
Coffee	Radio and television receivers	1.2
Legal services, short-run	Automobiles, short-run	1.2-1.5
Tobacco products, short-run	Restaurant meals	2.3
Residential natural gas, long-run	Airline travel, long-run	2.4
Fish (cod) consumed at home	Fresh green peas	2.8
Physician services	Foreign travel, long-run	4.0
Taxi, short-run	Chevrolet automobiles	4.0
Gasoline, long-run	Fresh tomatoes	4.6

From http://www.mackinac.org/archives/1997/s1997-04.pdf; cited sources: James D. Gwartney and Richard L. Stroup,, Economics: Private and Public Choice, 7th ed., 1995; 8th ed., 1997; Hendrick S. Houthakker and Lester D. Taylor, Consumer Demand in the United States, 1929–1970 (1966; Cambridge: Harvard University Press, 1970); Douglas R. Bohi, Analyzing Demand Behavior (Baltimore: Johns Hopkins University Press, 1981); Hsaing-tai Cheng and Oral Capps, Jr., "Demand for Fish," American Journal of Agricultural Economics, August 1988; and U.S. Department of Agriculture.

When demand is linear, $x(p) = a - bp$, the elasticity of demand has the form

$$\varepsilon = \mathrm{bpa} - \mathrm{bp} = \mathrm{p\ a\ b} - \mathrm{p}$$

This case is illustrated in Figure 3.1.

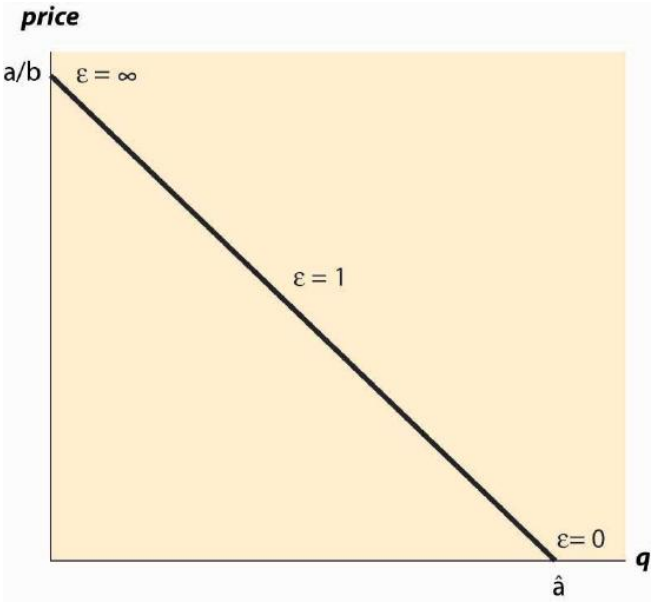


Figure 3.1 Elasticities for linear demand

If demand takes the form x(p) = a * $p^{-\varepsilon}$, then demand has **constant elasticity**, and the elasticity is equal to ε. In other words, the elasticity remains at the same level while the underlying variables (such as price and quantity) change.

The **elasticity of supply** is analogous to the elasticity of demand in that it is a unit-free measure of the responsiveness of supply to a price change, and is defined as the percentage increase in quantity supplied resulting from a small percentage increase in price. Formally, if s(p) gives the quantity supplied for each price p, the elasticity of supply, denoted by η (the Greek letter “eta,” chosen because epsilon was already taken) is

$$\eta = \mathrm{ds}\,\mathrm{sdpp} = \mathrm{psdsdp} = \mathrm{ps}'(\mathrm{p})\mathrm{s}(\mathrm{p})$$

Again, similar to demand, if supply takes the form s(p) = a * p^{η}, then supply has constant elasticity, and the elasticity is equal to η. A special case of this form is linear supply, which occurs when the elasticity equals one.

Key Takeaways

- The elasticity of demand is the percentage decrease in quantity that results from a small percentage increase in price, which is generally denoted with the Greek letter epsilon, ε.
- The percentage change of total revenue resulting from a 1% change in price is one minus the elasticity of demand.
- An elasticity of demand that is less than one is defined as an inelastic demand. In this case, increasing price increases total revenue.
- A price increase will decrease total revenue when the elasticity of demand is greater than one, which is defined as an elastic demand.
- The case of elasticity equal to one is called unitary elasticity, and total revenue is unchanged by a small price change.
- If demand takes the form x (p) = a * $p^{-\varepsilon}$, then demand has constant elasticity, and the elasticity is equal to ε.
- The elasticity of supply is defined as the percentage increase in quantity supplied resulting from a small percentage increase in price.
- If supply takes the form s (p) = a * p^{η}, then supply has constant elasticity, and the elasticity is equal to η.

EXERCISES

1. Suppose a consumer has a constant elasticity of demand ε, and demand is elastic $(\varepsilon > 1)$. Show that expenditure increases as price decreases.
2. Suppose a consumer has a constant elasticity of demand ε, and demand is inelastic $(\varepsilon < 1)$. What price makes expenditure the greatest?
3. For a consumer with constant elasticity of demand $(\varepsilon > 1)$, compute the consumer surplus.
4. For a producer with constant elasticity of supply, compute the producer profits.

This page titled 3.1: Elasticity is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

3.2: Supply and Demand Changes 3.2: Supply and Demand ChangesEdit Update titlecancel

Learning Objectives

1. What are the effects of changes in supply and demand on price and quantity?
2. What is a useful approximation of these changes?

When the price of a complement changes—what happens to the equilibrium price and quantity of the good? Such questions are answered by comparative statics, which are the changes in equilibrium variables when other things change. The use of the term "static" suggests that such changes are considered without respect to dynamic adjustment; instead, one just focuses on the changes in the equilibrium level. Elasticities will help us quantify these changes.

How much do the price and quantity traded change in response to a change in demand? We begin by considering the constant elasticity case, which allows us to draw conclusions for small changes for general demand functions. We will denote the demand function by $\mathrm{q_d(p) = a * p^{-\varepsilon}}$ and supply function by $\mathrm{q_s(p) = bp^n}$. The equilibrium price p* is determined at the point where the quantity supplied equals to the quantity demanded, or by the solution to the following equation:

$$qd\,(p^*) = qs\,(p^*)$$

Substituting the constant elasticity formulae,

$$ap * (-\varepsilon) = qd(p*) = qs(p*) = bp * (\eta).$$

Thus,

$$ab = p * \varepsilon + \eta$$

or

$$p^* = (ab)1\varepsilon + \eta$$

The quantity traded, q*, can be obtained from either supply or demand, and the price:

$$q^* = qs\,(p^*) = bp^*\eta = b(ab)\eta\varepsilon + \eta = a\eta\varepsilon + \eta b\varepsilon\varepsilon + \eta$$

There is one sense in which this gives an answer to the question of what happens when demand increases. An increase in demand, holding the elasticity constant, corresponds to an increase in the parameter a. Suppose we increase a by a fixed percentage, replacing a by $a(1 + \Delta)$. Then price goes up by the multiplicative factor $(1 + \Delta)1\varepsilon + \eta$ and the change in price, as a proportion of the price, is $\Delta p^* p^* = (1 + \Delta)1\varepsilon + \eta - 1$. Similarly, quantity rises by $\Delta q^* q^* = (1 + \Delta)\eta\varepsilon + \eta - 1$

These formulae are problematic for two reasons. First, they are specific to the case of constant elasticity. Second, they are moderately complicated. Both of these issues can be addressed by considering small changes—that is, a small value of Δ. We make use of a trick to simplify the formula. The trick is that, for small Δ,

$$(1 + \Delta)\mathbf{r} \approx 1 + r\Delta$$

The squiggly equals sign ≅ should be read, "approximately equal to."The more precise meaning of ≅ is that, as Δ gets small, the size of the error of the formula is small even relative to δ. That is, $(1 + \Delta)r \approx 1 + r\Delta$ means $(1 + \Delta)r - (1 + r\Delta)\Delta \to \Delta \to 00$. Applying this insight, we have the following:

For a small percentage increase Δ in demand, quantity rises by approximately ηΔ ε+η percent and price rises by approximately Δ ε+η percent.

The beauty of this claim is that it holds even when demand and supply do not have constant elasticities because the effect considered is local and, locally, the elasticity is approximately constant if the demand is "smooth."

Key Takeaways

- For a small percentage increase Δ in demand, quantity rises by approximately $\eta\Delta\varepsilon + \eta$ percent and price rises by approximately $\Delta\varepsilon + \eta$ percent.

- For a small percentage increase Δ in supply, quantity rises by approximately $\varepsilon\Delta\varepsilon + \eta$ percent and price falls by approximately $\Delta\varepsilon + \eta$ percent.

EXERCISES

1. Show that, for a small percentage increase Δ in supply, quantity rises by approximately εΔ ε+η percent and price falls by approximately Δ ε+η percent.
2. If demand is perfectly inelastic (ε = 0), what is the effect of a decrease in supply? Apply the formula and then graph the solution.
3. Suppose demand and supply have constant elasticity equal to 3. What happens to equilibrium price and quantity when the demand increases by 3% and the supply decreases by 3%?
4. Show that elasticity can be expressed as a constant times the change in the log of quantity divided by the change in the log of price (i.e., show $\varepsilon = Adlnx(p)dlnp\backslash]$). Find the constant A.
5. A car manufacturing company employs 100 workers and has two factories, one that produces sedans and one that makes trucks. With m workers, the sedan factory can make m^2 sedans per day. With n workers, the truck factory can make $5n^3$ trucks per day. Graph the production possibilities frontier.
6. In Exercise 5, assume that sedans sell for $20,000 and trucks sell for $25,000. What assignment of workers maximizes revenue?

This page titled 3.2: Supply and Demand Changes is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

4: The U.S. Economy

4.1: Basic Demographics
4.2: Education
4.3: Households and Consumption
4.4: Production
4.5: Government
4.6: Trade
4.7: Fluctuations

This page titled 4: The U.S. Economy is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

4.1: Basic Demographics 4.1: Basic DemographicsEdit Update titlecancel

An important aspect of economics is economic statistics, and an army of economists collects and analyzes these statistics. This chapter presents an overview of the economic activity of the United States. How much do you need to know about these statistics? It would be ridiculous to memorize them. At the same time, it would be undesirable to be ignorant of how we are changing and how we are not.I apologize to those using the book in foreign countries; this chapter is about the United States not because it is more important but because I know it better. Encourage your professor to write a chapter on your country! All of the statistics in this chapter come from Fedstats, www.fedstats.gov/, from FRED, http://research.stlouisfed.org/fred2/, and from the NBER, http://www.nber.org/.

Learning Objectives

- Who lives in the United States?

There are about 300 million people in the United States, up from 76 million in 1900.

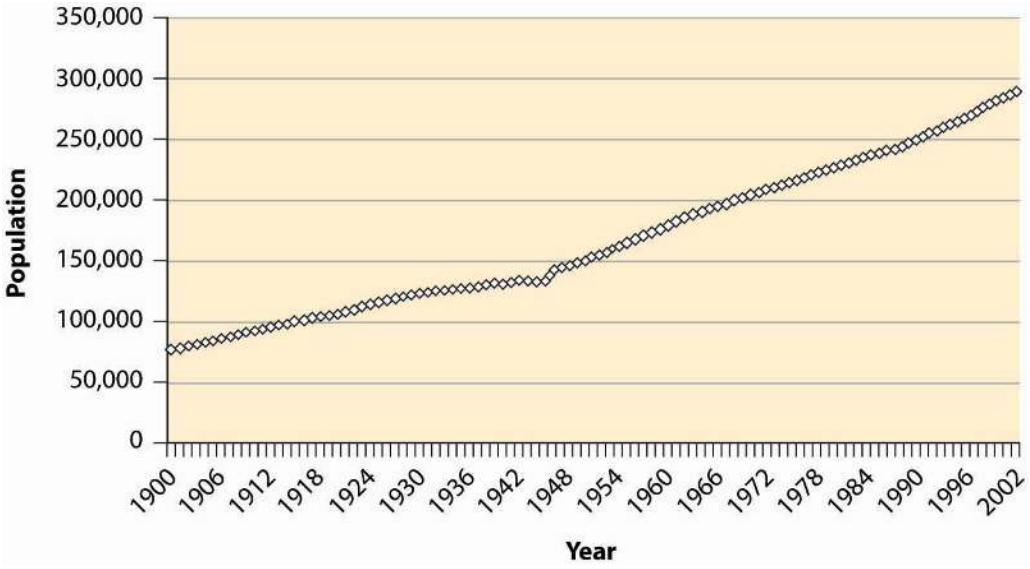


Figure 4.1 U.S. resident population

During the last century, the U.S. population has become primarily an urban population, growing from 40% to 80% urban. The population is primarily white, with 12%–13% African American and 4% classified as other. These proportions are relatively stable over the century, with the white population falling from 89% to 83%. The census is thought to understate minority populations because of greater difficulties in contacting minorities. The census does not attempt to classify people but instead accepts people's descriptions of their own race.

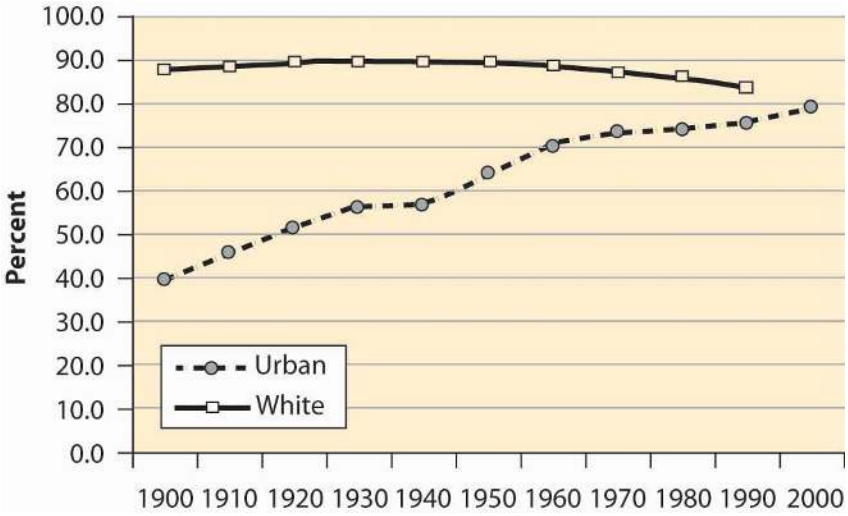


Figure 4.2 U.S. urban and white population

The U.S. population has been aging significantly, with the proportion of seniors (over 65 years of age) tripling over the past century, and the proportion of young people dropping by over one-third. Indeed, the proportion of children between 0 and 5 years old has dropped from 12.1% of the population to under 7%.

Figure 4.3 Population proportions by age group

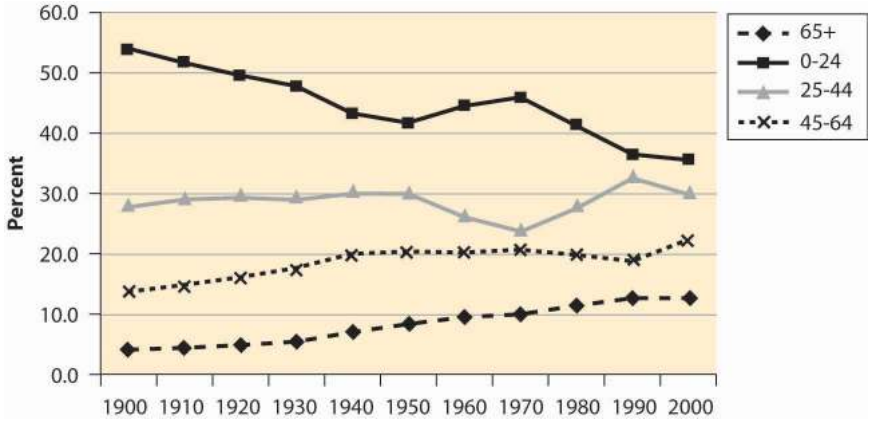


Figure 4.4.

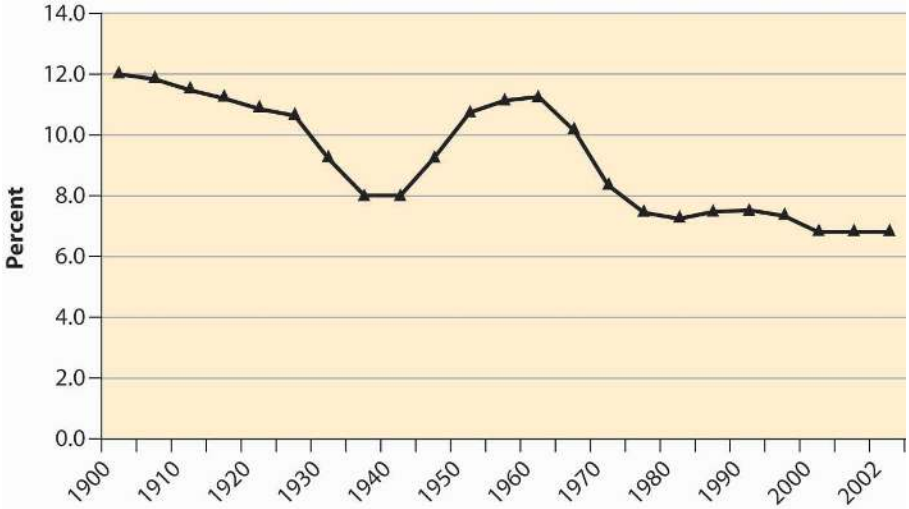


Figure 4.4 Proportion of population under age 5

The aging of the American population is a consequence of greater life expectancy. When social security was created in 1935, the average American male lived to be slightly less than 60 years old. The social security benefits, which didn't start until age 65, thus were not being paid to a substantial portion of the population.

Figure 4.5 U.S. life expectancy at birth

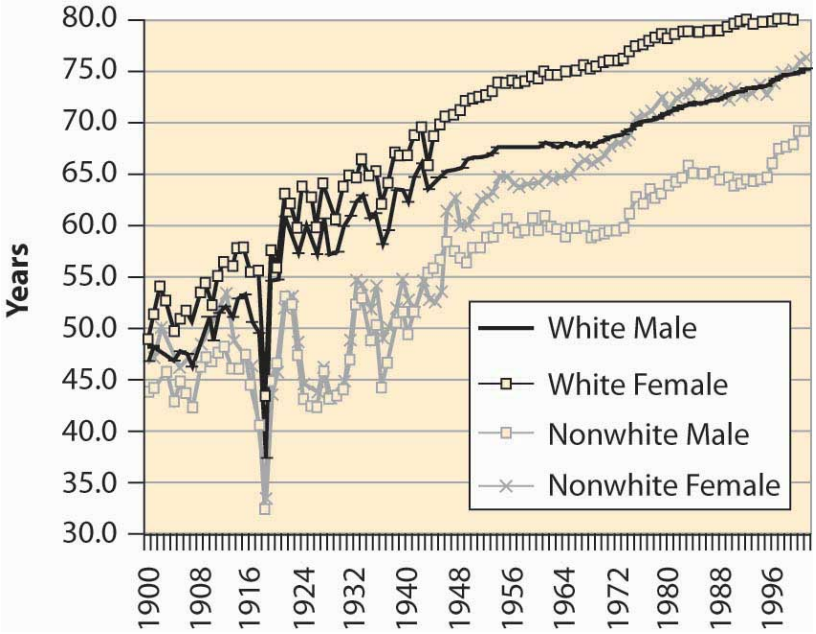


Figure 4.5 shows life expectancy at birth, thus including infant mortality. The significant drop in life expectancy in 1918—to nearly 30 years old for nonwhites—is primarily a consequence of the great influenza, which killed about 2.5% of the people who contracted it and killed more Americans in 1918 than did World War I. The Great Depression (1929–1939) also reduced life expectancy. The steady increase in life expectancy is also visible, with white females now living 80 years on average.

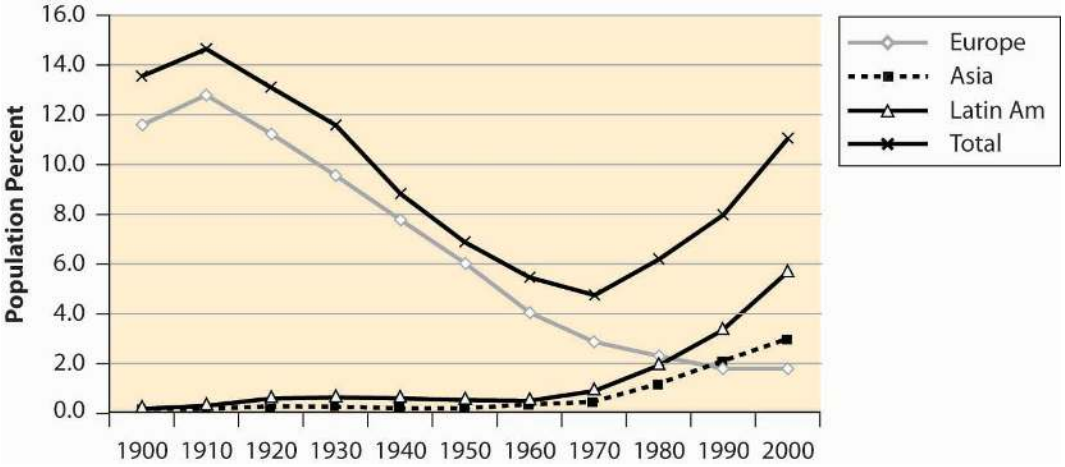


Figure 4.6 U.S. immigrant population (percentages) by continent of origin

It is said that the United States is a country of immigrants, and a large fraction of the population had ancestors who came from elsewhere. Immigration into this United States, however, has been increasing after a long decline, and the fraction of the population that was born in foreign countries is about 11%—one in nine.

Figure 4.7 National origin of immigrants, 1900–2000

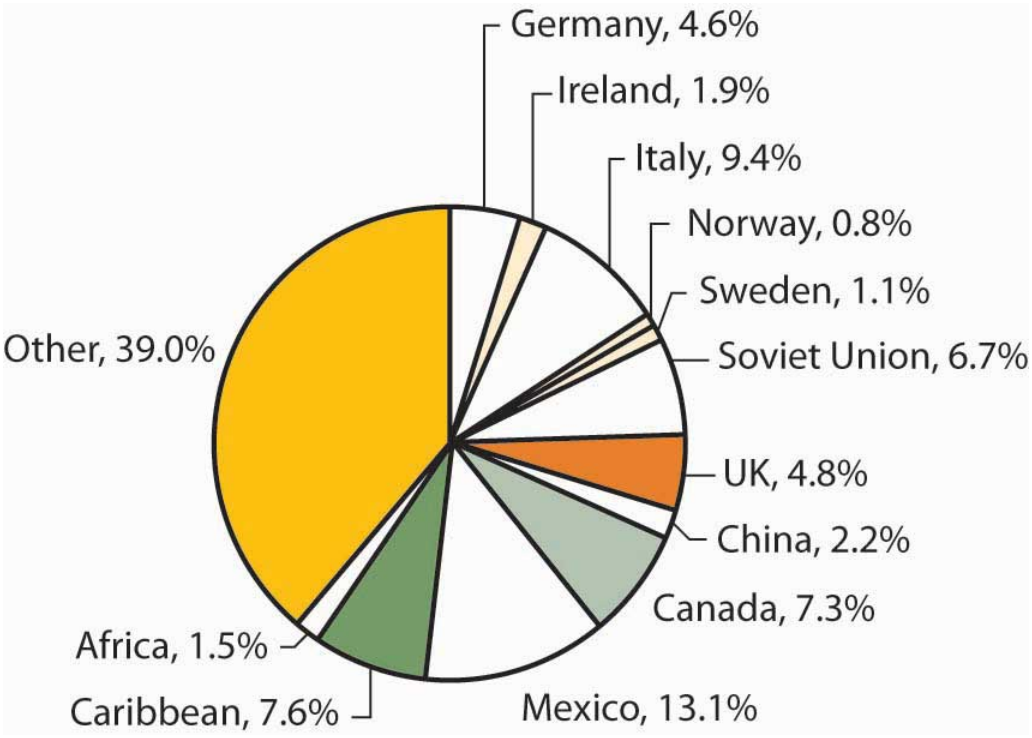


Figure 4.7 aggregates the total country-of-origin data over the century to identify the major sources of immigrants.

One hears a lot about divorce rates in the United States, with statements like "Fifty percent of all marriages end in divorce." Although it has grown, the divorced population is actually a small fraction of the population of the United States.

Figure 4.8 Male marital status (percentages)

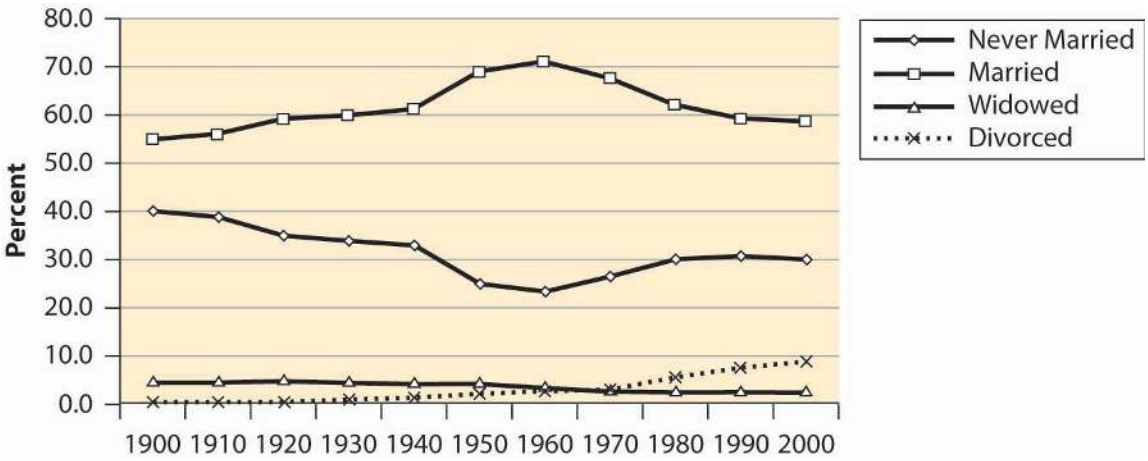
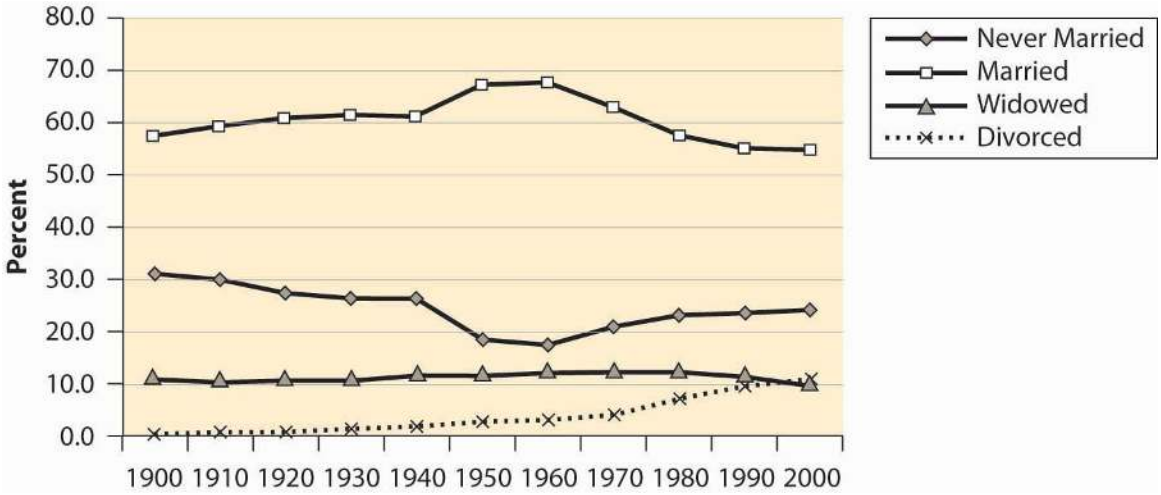


Figure 4.9 Female marital status (percentages)



Marriage rates have fallen, but primarily because the “never married” category has grown. Some of the “never married” probably represent unmarried couples, since the proportion of children from unmarried women has risen fairly dramatically. Even so, marriage rates are greater than they were a century ago. However, a century ago there were more unrecorded and common-law marriages than there probably are today.

Figure 4.10 Births to unwed mothers (percentages)

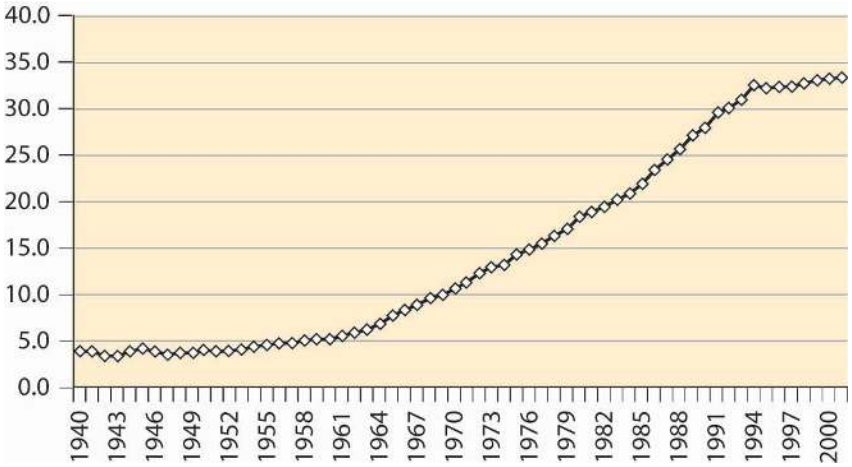


Figure 4.11.

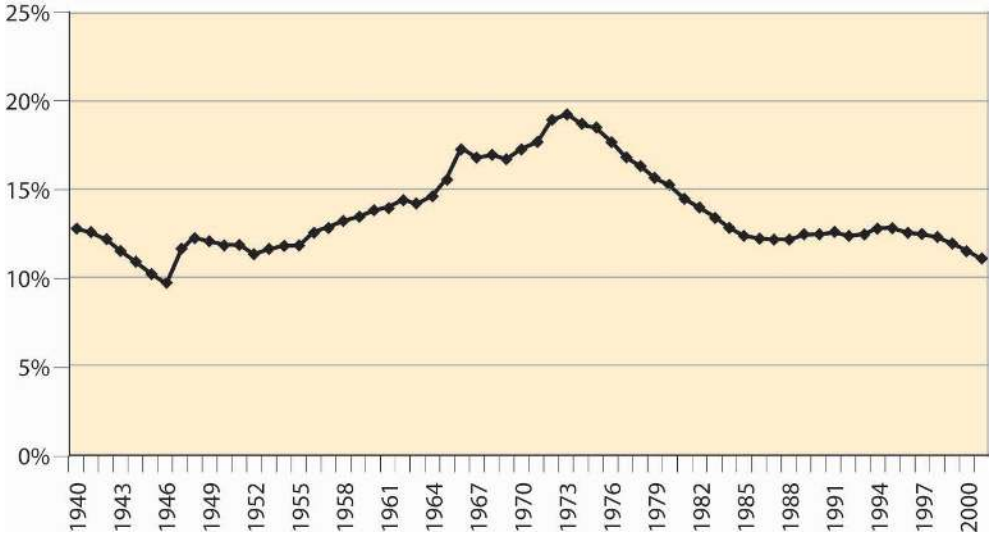


Figure 4.11 Births to women age 19 or less (percentages)

Key Takeaways

- No one in his or her right mind memorizes the takeaways of this chapter; the goal is to have a sense of one's nation.
- There are about 300 million people in the United States, up from 76 million in 1900.
- The U.S. population has become primarily an urban population, growing from 40% to 80% urban in the past century.
- The population is primarily white, with 12%–13% African American.
- The U.S. population has aged, with the proportion of seniors (over 65 years of age) tripling over the past century, and the proportion of young people dropping by over one-third.
- The baby boom was a dramatic increase in births for the years 1946 to 1964.
- The aging of the American population is a consequence of greater life expectancy.
- About 11% of Americans were born in foreign countries.
- The divorced population is about 10%.
- Marriage rates have fallen, but primarily because the "never married" category has grown.

This page titled 4.1: Basic Demographics is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

4.2: Education 4.2: EducationEdit Update titlecancel

Learning Objectives

- Who goes to school and how much?

Why are the Western nations rich and many other nations poor? What creates the wealth of the developed nations? Modern economic analysis attributes much of the growth of the United States and other developed nations to its educated workforce, and not to natural resources. Japan, with a relative scarcity of natural resources but a highly educated workforce, is substantially richer than Brazil, with its abundance of natural resources.

Figure 4.12 Educational attainment in years (percentage of population)

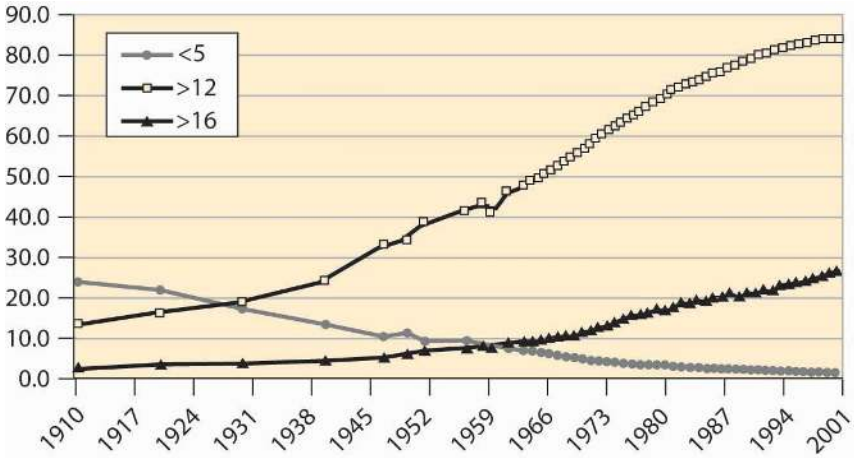


Figure 4.12. Slightly fewer women (25% vs. 28%)complete 4 years of university, although women are more likely to complete 4 years of high school.

Graduation rates are somewhat below the number of years completed, so that slightly less than three quarters of the U.S. population actually obtain their high school degree. Of those obtaining a high school degree, nearly half obtain a university or college degree.

Figure 4.13 Graduation rates (percentages)

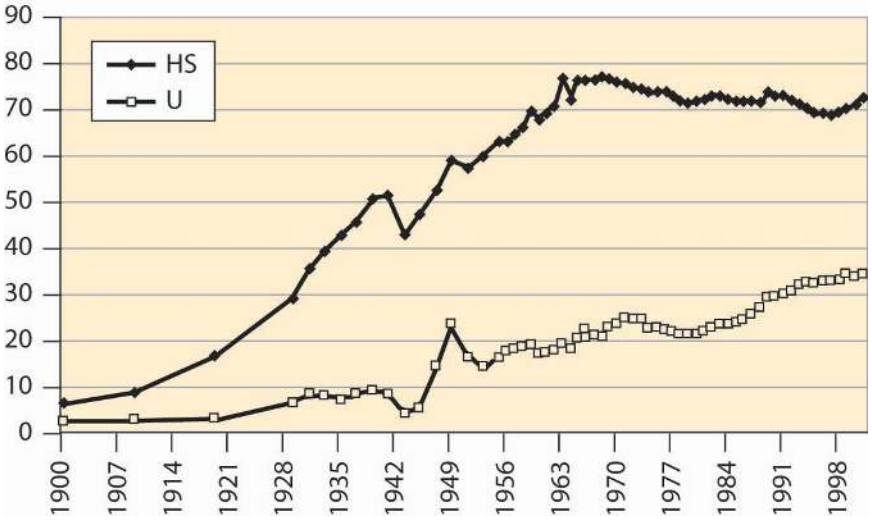


Figure 4.13. First, high school completion dropped significantly during World War II (1940–1945) but rebounded afterward. Second, after World War II, college graduation spiked when many U.S. soldiers were sent to university by the government under a program called the “GI Bill.”The etymology of GI, as slang for U.S. soldiers, is disputed, with candidates including “Government Issue,” “General Infantry,” and “Galvanized Iron”—the latter a reference to trash cans that looked like German World War I artillery.

As the number of high school students rose, the portion of high school graduates going to university fell, meaning that a larger segment of the population became high school educated. This increase represents the creation of the U.S. middle class; previously, high school completion and university attendance was in large part a sign of wealth. The creation of a large segment of the population who graduated from high school but didn't attend university led to a population with substantial skills and abilities but no inherited wealth and they became the middle class.

High school completion has been declining for 30 years. This is surprising given the high rate of financial return to education in the United States. Much of the reduction in completion can be attributed to an increase in General Education Development (GED) certification, which is a program that grants diplomas (often erroneously thought to be a "General Equivalent Degree") after successfully passing examinations in five subject areas. Unfortunately, those people who obtain GED certification are not as successful as high school graduates, even marginal graduates, and indeed the GED certification does not seem to help students succeed, in comparison with high school graduation.In performing this kind of analysis, economists are very concerned with adjusting for the type of person. Smarter people are more likely to graduate from high school, but one doesn't automatically become smart by attending high school. Thus, care has been taken to hold constant innate abilities, measured by various measures like IQ scores and performance on tests, so that the comparison is between similar individuals, some of whom persevere to finish school, some of whom don't. Indeed, some studies use identical twins.

Key Takeaways

- An estimated 85% of the U.S. population completes 12 years of schooling, not counting kindergarten.
- One quarter of the population completes at least 4 years of university.
- High school graduates comprise the bulk of the middle class.
- High school completion has been declining for 30 years. This is surprising given the high rate of financial return to education in the United States.

This page titled 4.2: Education is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

4.3: Households and Consumption 4.3: Households and ConsumptionEdit Update titlecancel

Learning Objectives

- How much stuff do we have?

There are approximately 100 million **households**—a group of people sharing living quarters—in the United States. The number of residents per household has consistently shrunk during this century, from over four to under three, as illustrated in Figure 4.14.

Figure 4.14 Household occupancy

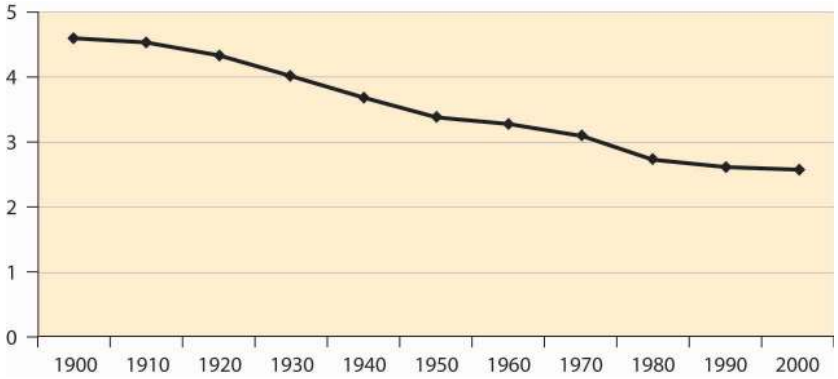


Figure 4.15. More women live alone than men, even though four times as many families with a single adult member are headed by women. This discrepancy—many more women both living on their own and living with children and no partner, even though there are about the same number of men and women born—is accounted for by the greater female longevity already noted.

Figure 4.15 Proportion of households by type

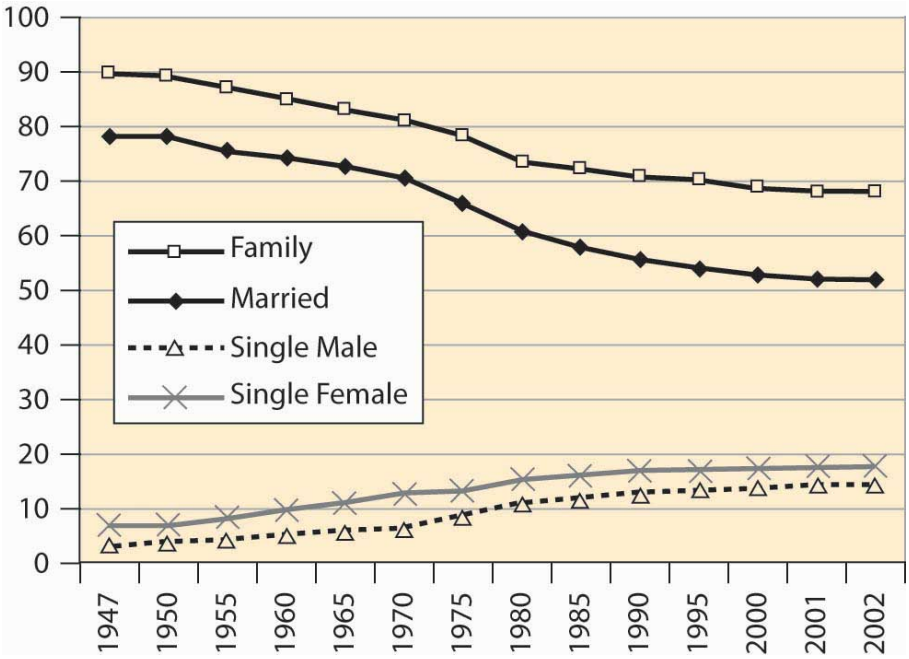


Figure 4.16. This represents a fourfold increase over 1925 to 1975.

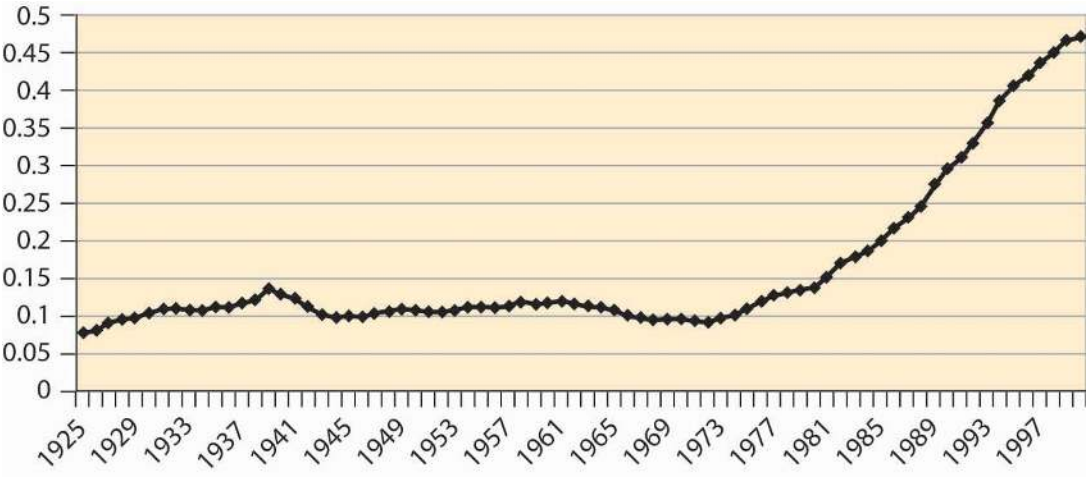


Figure 4.16 Percentage of incarcerated residents

Ten percent of households do not have an automobile, and 97.6% have a telephone. So-called land line telephones may start to fall as apartment dwellers, especially students, begin to rely exclusively on cell phones. Just under 99% of households have complete plumbing facilities (running water, bath or shower, flush toilet), up from 54.7% in 1940.

How much income do these households make? What is the distribution of income? One way of assessing the distribution is to use quintiles to measure dispersion. A **quintile** is one fifth, or 20%, of a group. Thus the top income quintile represents the top 20% of income earners, the next represents those ranking 60%–80%, and so on. Figure 4.17 shows the earnings of the top, middle, and bottom quintiles.

Figure 4.17 Income shares for three quintiles

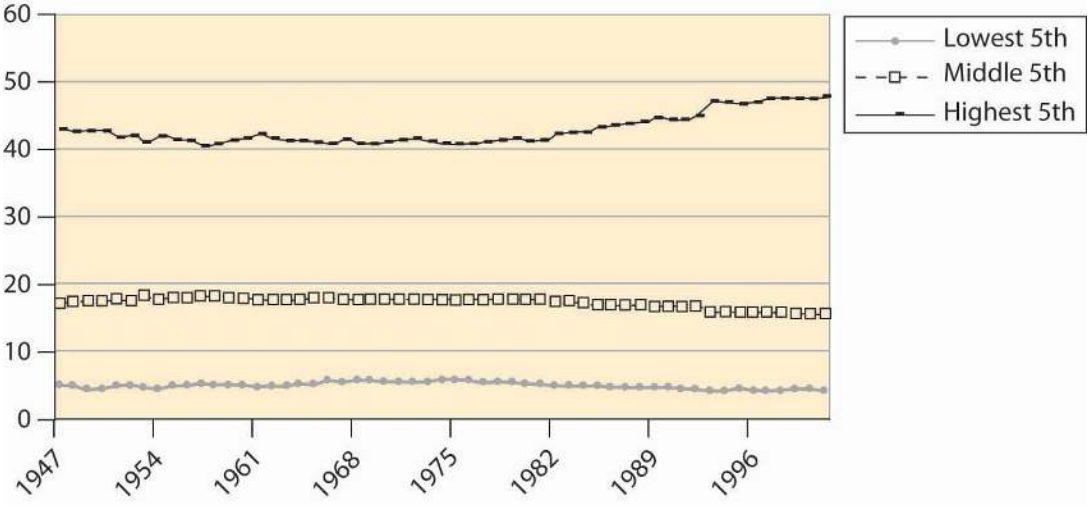
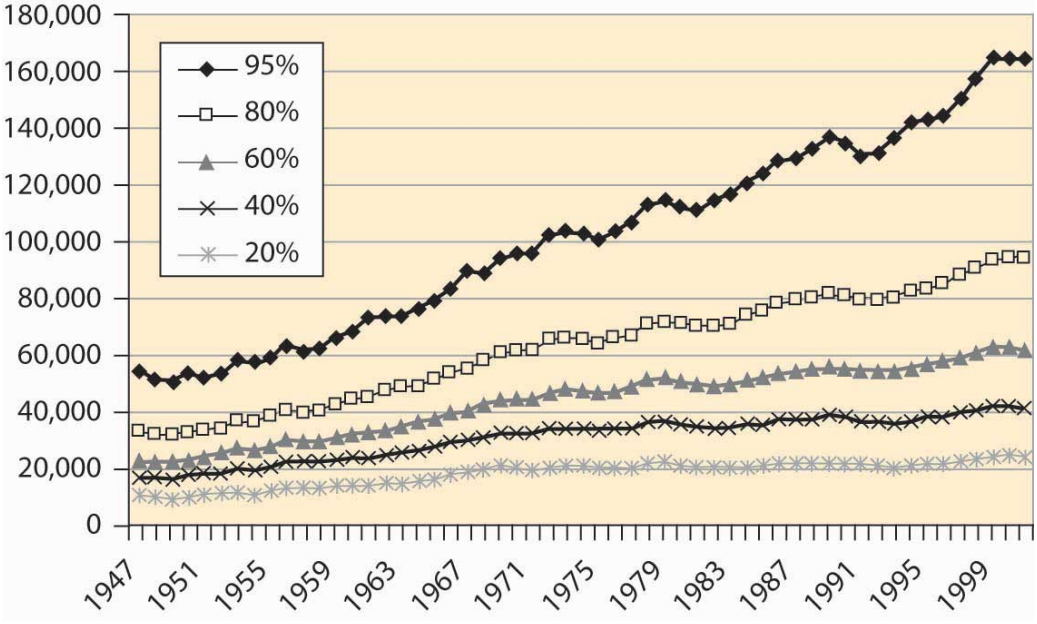


Figure 4.18 Family income



Figures like these suggest that families are getting poorer, except for an elite few. However, families are getting richer, just not as fast as the top quintile.

Figure 4.19 Family income, cumulative percentage change

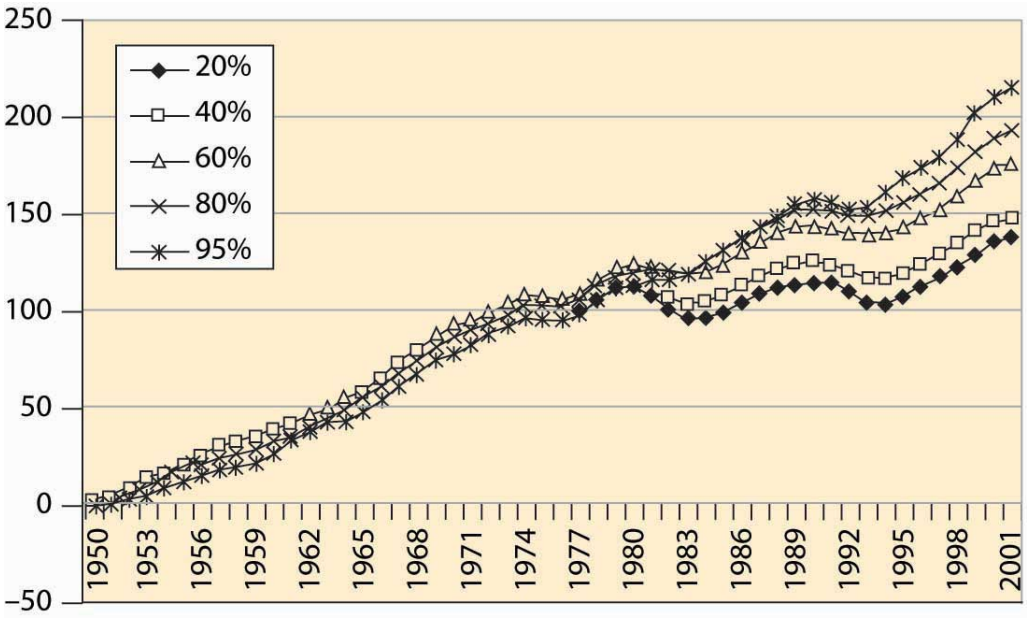


Figure 4.19.

Real income gains in percentage terms have been larger for richer groups, even though the poor have also seen substantially increased incomes.

If the poor have fared less well than the rich in percentage terms, how have African Americans fared? After World War II, African American families earned about 50% of white family income. This ratio has risen gradually, noticeably in the 1960s after the **Civil Rights Act**—legislation that prohibited segregation based on race in schools, public places, and employment—that is credited with integrating workplaces throughout the southern United States. African American family income lagged white income growth throughout the 1980s but has been rising again, a trend illustrated in Figure 4.20.

Figure 4.20 Black family income as a percentage of white income

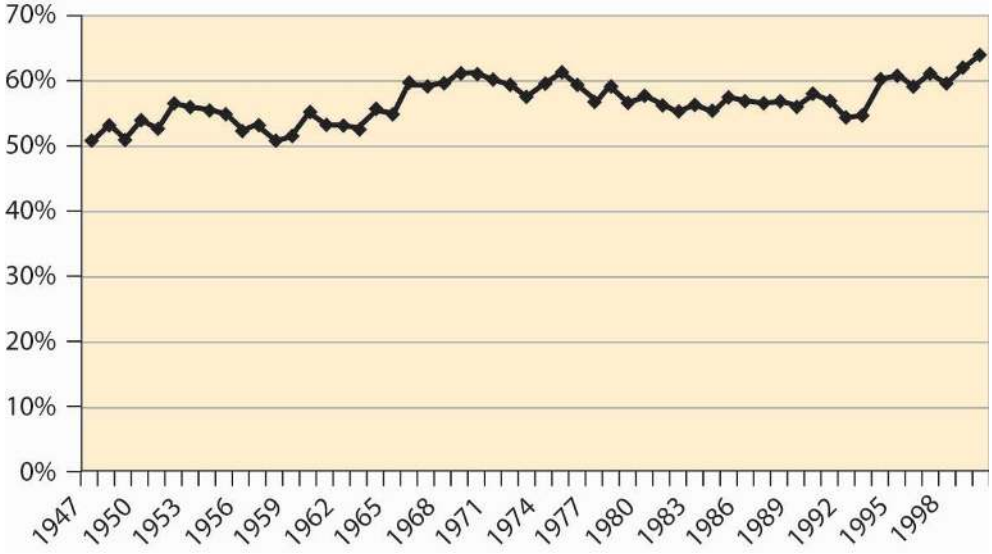


Figure 4.21 shows the CPI over most of the past century, where 1982 is set as the reference year.

There have been three major inflations in the past century. Both World War I and World War II, with a large portion of the goods and services diverted to military use, saw significant inflations. In addition, there was a substantial inflation during the 1970s, after the Vietnam War in the 1960s. The price level fell during the **Great Depression**, a prolonged and severe economic downturn from 1929 to 1939. Falling price levels create investment problems because inflation-adjusted interest rates, which must adjust for deflation, are forced to be high, since unadjusted interest rates cannot be negative. Changes in the absolute price level are hard to estimate, so the change is separately graphed in Figure 4.22.

Figure 4.21 Consumer price index (1982 = 100)

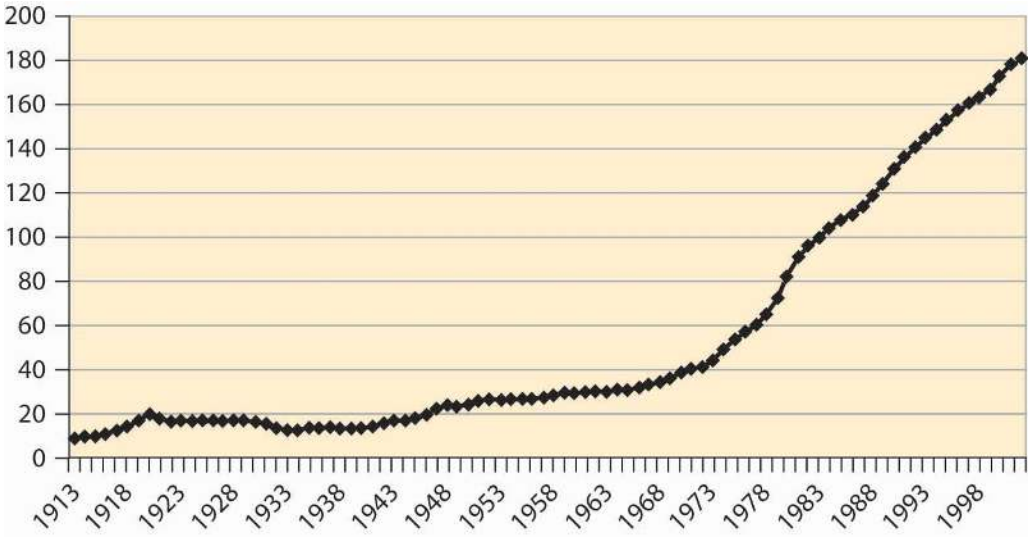


Figure 4.22 CPI percentage changes

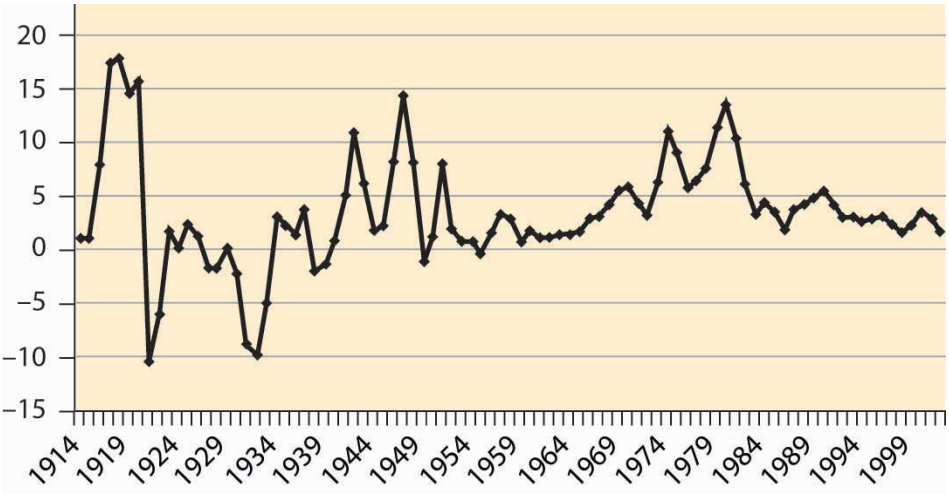


Figure 4.23 shows that the percentage of pre-tax household income spent on food has fallen from 25% to about 10%. This fall is a reflection of greater incomes and the fact that the real cost of food has fallen.

Moreover, a much greater fraction of expenditures on food are spent away from home, a fraction that has risen from under 15% to 40%.

Figure 4.23 Food expenditure as percentage of income, and proportion spent out

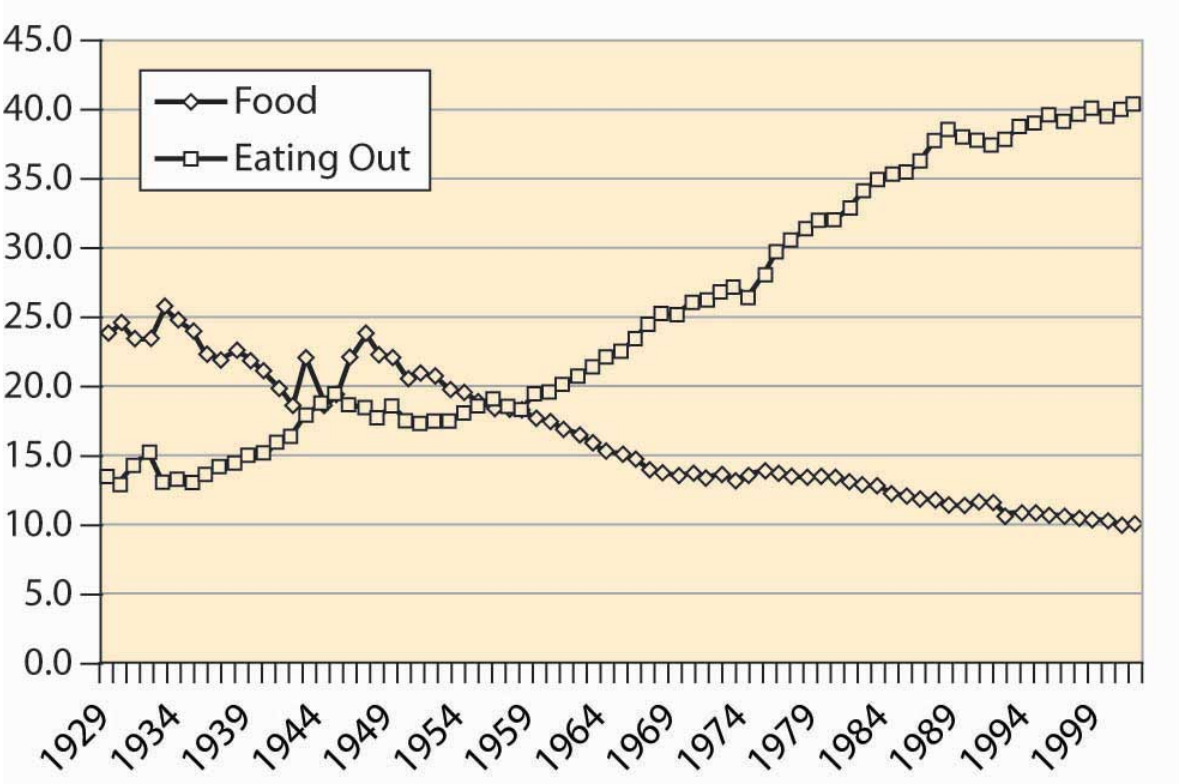


Figure 4.24.

Figure 4.24 After-tax income shares

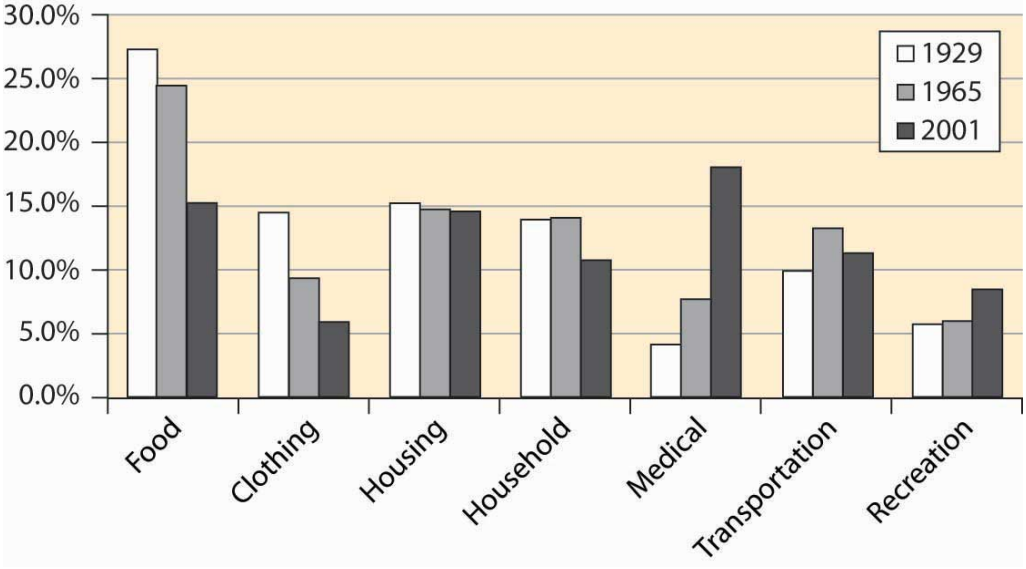


Figure 4.23 because these figures use after-tax disposable income, rather than pre-tax income.) This is not only a consequence of our aging population but also of the increased technology available.

Key Takeaways

- There are approximately 100 million households in the United States.
- The number of residents per household has shrunk, from over four to under three, over the past 100 years.
- About 60% of households live in single-family detached homes.
- Slightly less than 0.5% of the population is incarcerated in state and federal prisons. This represents a four-fold increase over 1925 to 1975.
- Ten percent of households do not have an automobile, and 97.6% have a telephone.
- Just under 99% of households have complete plumbing facilities (running water, bath or shower, flush toilet), up from 55% in 1940.
- A quintile (or fifth) is a group of size 20%.
- The earnings of the top quintile fell slightly until the late 1960s, when it began to rise. All other quintiles lost income share to the top quintile starting in the mid-1980s. Figures like these suggest that families are getting poorer, except for an elite few. However, families are getting richer, just not as fast as the top quintile.
- Just after World War II, African American families earned about 50% of white family income. This ratio has risen gradually, noticeably in the 1960s after the 1964 Civil Rights Act.
- The consumer price index (CPI), which adjusts for what it costs to buy a “standard” bundle of food, clothing, housing, electricity, and other items, is the most common price index.
- There have been three major inflations in the past century, associated with World War I, World War II, and the 1970s. The price level fell during the Great Depression (1929–1939).
- The cost of food has fallen quite dramatically over the past century.

EXERCISES

1. Have prices actually risen? Economists generally agree that the meaning of “prices have risen” is that you would prefer past prices to current prices. What makes this challenging is that the set of available products change over time. Cars have gone up significantly in price but are also more reliable. Would you be better off with your current income in 1913 than today? You would be very rich with current average income in 1913 but would not have access to modern medicine, television, electronics, refrigeration, highways, and many other technologies. If you made $40,000 annually in 1913, how would you live and what would you buy? (Do some research.)
2. Compare a $40,000 income in 1980 to the present. What differences are there in available products? In the quality of products? How rich does $40,000 make you in each time period? In which period would you choose to live, and why?

This page titled 4.3: Households and Consumption is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

4.4: Production 4.4: ProductionEdit Update titlecancel

Learning Objectives

- What do we make?

We learned something about where we live and what we buy. Where do we get our income? Primarily, we earn by providing goods and services. Nationally, we produce about $11 trillion worth of goods and services. Broadly speaking, we spend that $11 trillion on personal consumption of goods and services, savings, and government. This, by the way, is often expressed as Y = C + I + G, which states that income (Y) is spent on consumption (C), investment (I, which comes from savings), and government (G). One can consume imports as well, so the short-term constraint looks like Y + M = C + I + G + X, where M is imports and X is exports.

How much does the United States produce? Economists measure output with the **gross domestic product (GDP)**, which is the value of traded goods and services produced within the borders of the United States. GDP thus excludes output of Japanese factories owned by Americans but includes the output of U.S. factories owned by the Japanese.

Importantly, GDP excludes nontraded goods and services. Thus, unpaid housework is not included. If you clean your own home, and your neighbor cleans his or her home, the cleaning does not contribute to GDP. On the other hand, if you and your neighbor pay each other to clean each other's homes, GDP goes up by the payments, even though the actual production of goods and services remains unchanged. Thus, GDP does not measure our total output as a nation, because it neglects unpaid services. Why does it neglect unpaid services? Primarily because we can't readily measure them. Data on transactions are generated by tax information and reporting requirements imposed on businesses. For the same reason, GDP neglects illegal activities as well, such as illegal drug sales and pirated music sales. Thus, GDP is not a perfect measure of the production of our society. It is just the best measure we have.

Figure 4.25 shows the growth in GDP and its components of personal consumption, government expenditures, and investment. The figures are expressed in constant 1996 dollars—that is, adjusted for inflation. The figure for government includes the government's purchases of goods and services—weapons, highways, rockets, pencils—but does not include transfer payments like social security and welfare programs. Transfer payments are excluded from this calculation because actual dollars are spent by the recipient, not by the government. The cost of making the transfer payments (e.g., printing and mailing the checks), however, is included in the cost of government.

Figure 4.25 Output, consumption, investment, and government

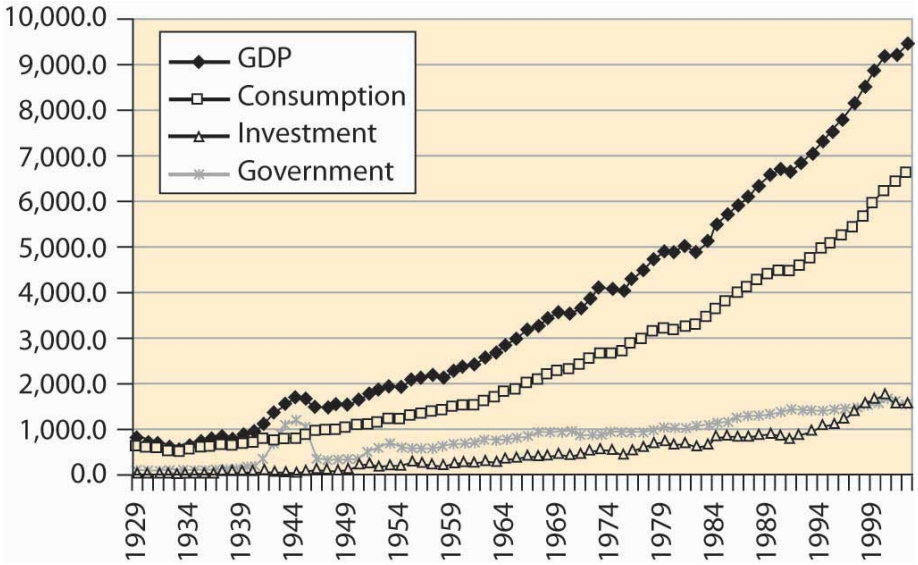


Figure 4.25, because economic activity is growing at a constant percentage. Consequently, economists often use a **logarithmic scale** rather than a dollar scale. A logarithmic scale has the useful property that a straight line gives constant percentage growth. Consider a variable x that takes on values x_t at time t. Define $\%\Delta x$ to be the percentage change:

$$\%\Delta x = xt - xt - 1 \times t - 1$$

https://socialsci.libretexts.org/@go/page/43762

Then

$$\log(xt) = \log(xt-1) + \log(xtxt-1) = \log(xt-1) + \log(1+\%\Delta x)$$

Thus, if the percentage change is constant over time, log(x_t) will be a straight line over time. Moreover, for small percentage changes,

$$\log(1+\%\Delta x) \approx \%\Delta x$$

so that the slope is approximately the growth rate.The meaning of ≈ throughout this book is "to the first order." Here that means lim %Δx→0 log(1+%Δx)−%Δx %Δx =0. Moreover, in this case the errors of the approximation are modest up to about 25% changes. Figure 4.26 shows these statistics with a logarithmic scale.

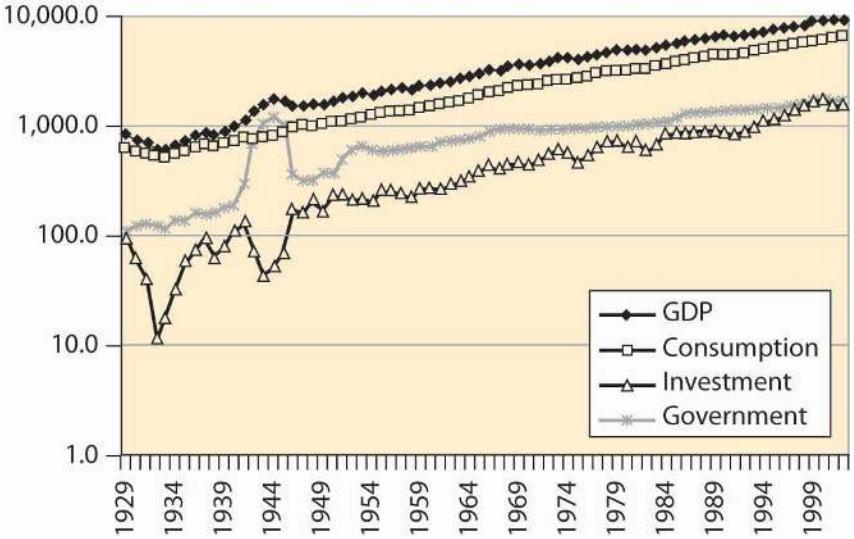


Figure 4.26 Major GDP components in log scale

Immediately noticeable is the approximately constant growth rate from 1950 to the present, because a straight line with a log scale represents a constant growth rate. In addition, government has grown much more slowly (although recall that transfer payments, another aspect of government, aren't shown). A third feature is the volatility of investment—it shows much greater changes than output and consumption. Indeed, during the Great Depression (1929–1939), income fell somewhat, consumption fell less, government was approximately flat, and investment plunged to 10% of its former level.

Some of the growth in the American economy has arisen because there are more of us. Double the number of people, and consume twice as many goods, and individually we aren't better off. How much are we producing per capita, and how much are we consuming?

U.S. output of goods, services, and consumption has grown substantially over the past 75 years, a fact illustrated in Figure 4.27. In addition, consumption has been a steady percentage of income. This is more clearly visible when income shares are plotted in Figure 4.28.

Figure 4.27 Per capita income and consumption

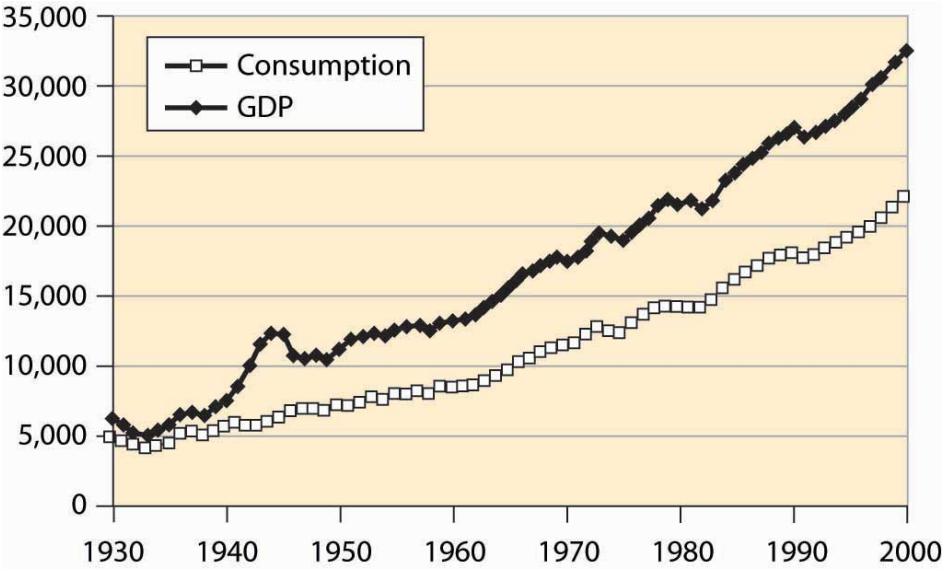
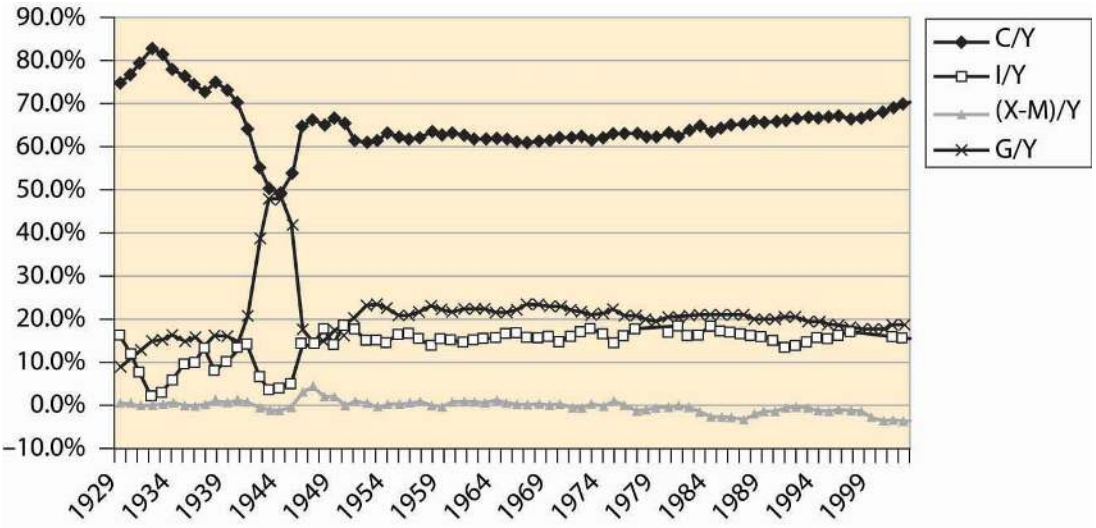


Figure 4.28 Consumption, investment, and government (% GDP)



Consumption was a very high portion of income during the Great Depression because income itself fell. Little investment took place. The wartime economy of World War II reduced consumption to below 50% of output, with government spending a similar fraction as home consumers. Otherwise, consumption has been a relatively stable 60%–70% of income, rising modestly during the past 20 years, as the share of government shrank and net imports grew. Net imports rose to 4% of GDP in 2001.

The most basic output of our economic system is food, and the U.S. economy does a remarkable job producing food. The United States has about 941 million acres under cultivation to produce food, which represents 41.5% of the surface area of the United States. Land use for agriculture peaked in 1952, at 1,206 million acres, and has been dwindling ever since, especially in the northeast where farms are being returned to forest through disuse. Figure 4.29 shows the output of agricultural products in the United States, adjusted to 1982 prices.

Figure 4.29 U.S. agricultural output, 1982 constant dollars

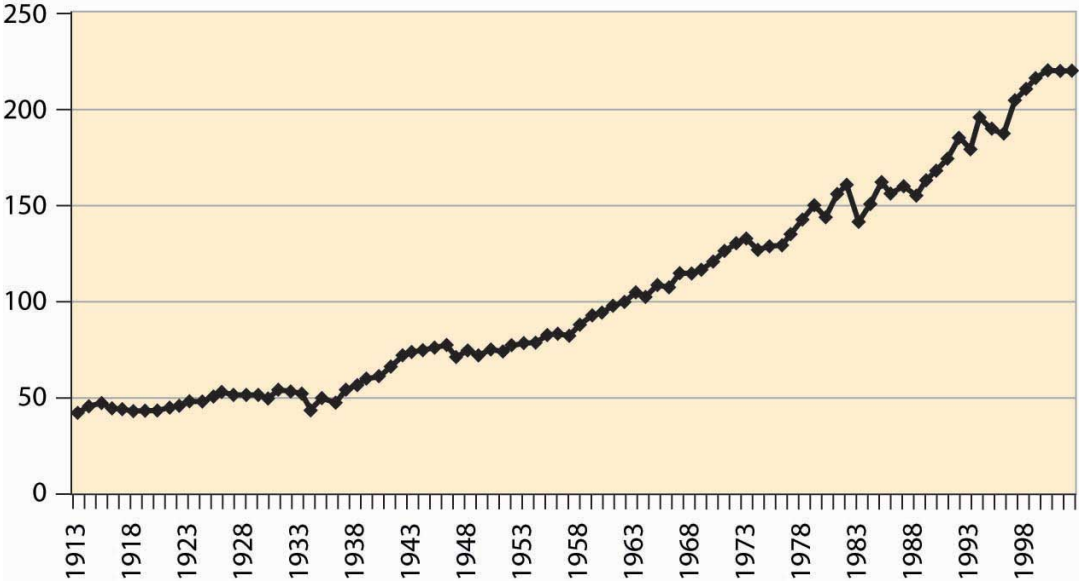


Figure 4.30.

Figure 4.30 Agricultural output, total and per worker (1982 dollars, log scale)

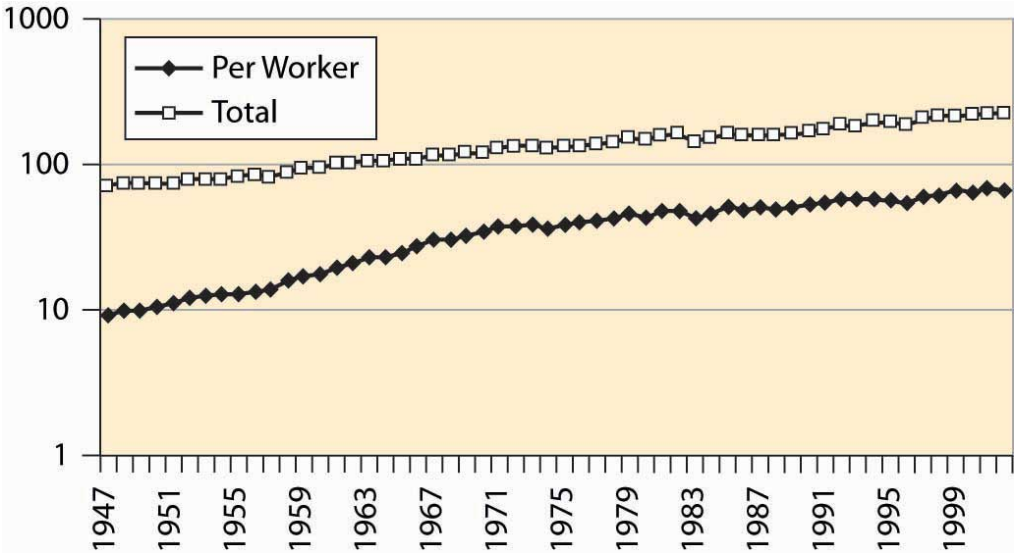


Figure 4.31 describes the major sectors of the U.S. economy. Because the data come from firms, agriculture is excluded, although goods and services provided to farms would be included.

https://socialsci.libretexts.org/@go/page/43762

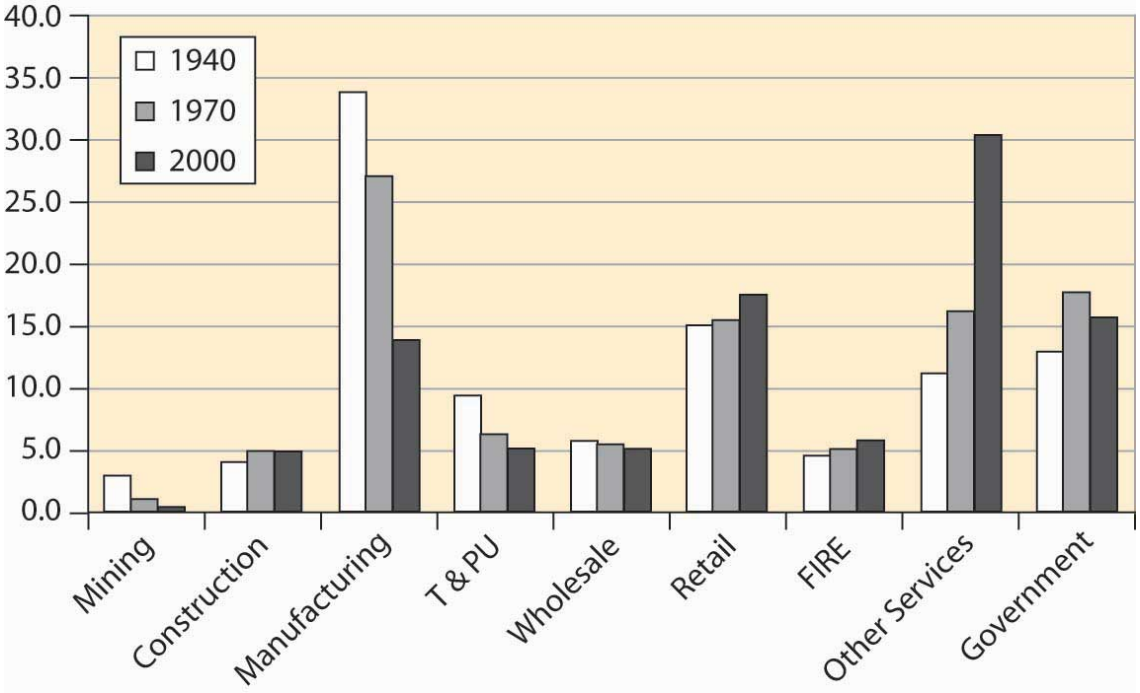


Figure 4.31 Major nonagricultural sectors of U.S. economy (% GDP)

Mining has diminished as a major factor in the U.S. economy, a consequence of the growth of other sectors and the reduction in the prices for raw materials. Contrary to many popular predictions, the prices of raw materials have fallen even as output and population have grown. We will see later in this book that the fall in prices of raw materials—ostensibly in fixed supply given the limited capacity of the earth—means that people expect a relative future abundance, either because of technological improvements in their use or because of large as yet undiscovered pools of the resources. An example of technological improvements is the substitution of fiber optic cable for copper wires. An enormous amount of copper has been recovered from telephone lines, and we can have more telephone lines and use less copper than was used in the past.

Manufacturing has become less important for several reasons. Many manufactured goods cost less, pulling down the overall value. In addition, we import more manufactured goods than in the past. We produce more services. T&PU stands for transportation and public utilities, and includes electricity and telephone services and transportation including rail and air travel. This sector has shrunk as a portion of the entire economy, although the components have grown in absolute terms. For example, the number of airplane trips has grown dramatically, as illustrated in Figure 4.32.

Figure 4.32 Air travel per capita

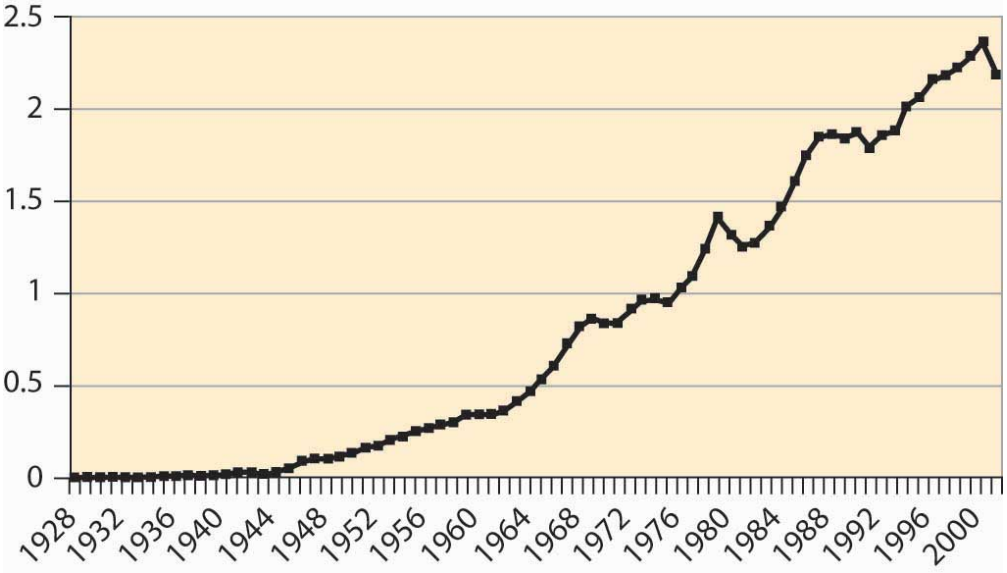


Figure 4.33 shows.

Figure 4.33 Electricity production (M kwh)

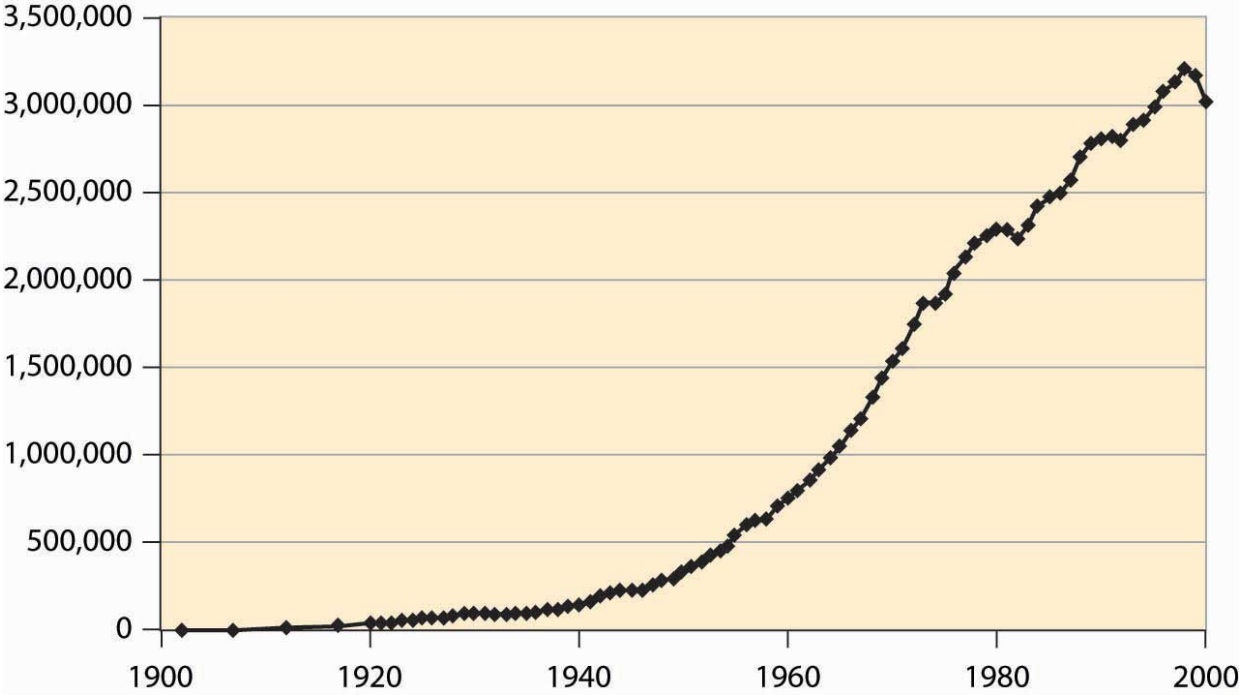


Figure 4.34.

Figure 4.34 Energy use (quadrillion BTUs)

https://socialsci.libretexts.org/@go/page/43762

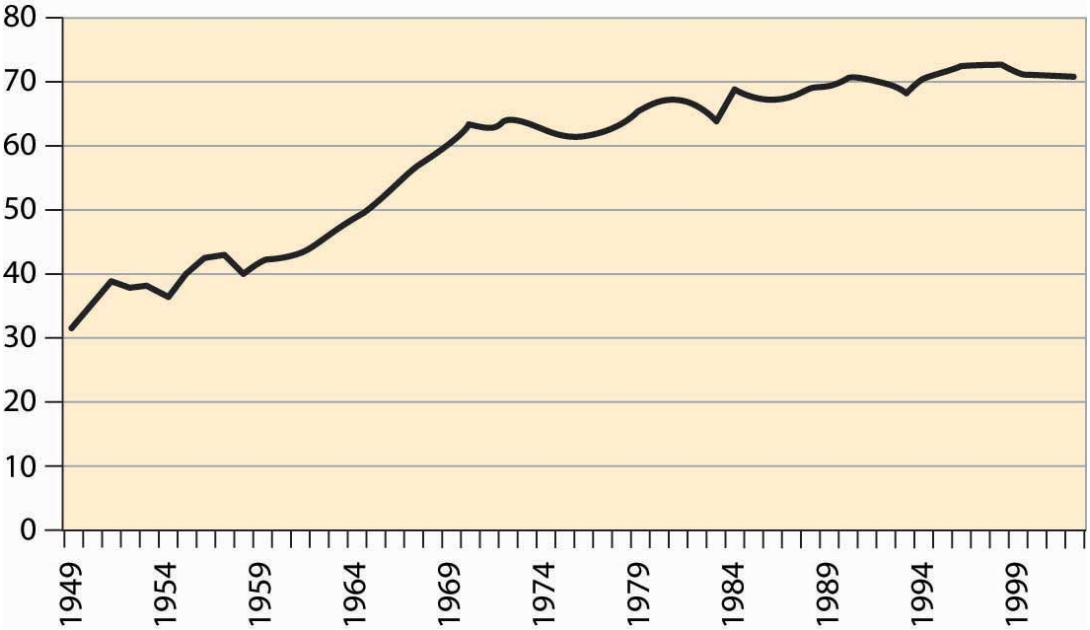


Figure 4.35, with miles on the left axis and cars per thousand on the right.

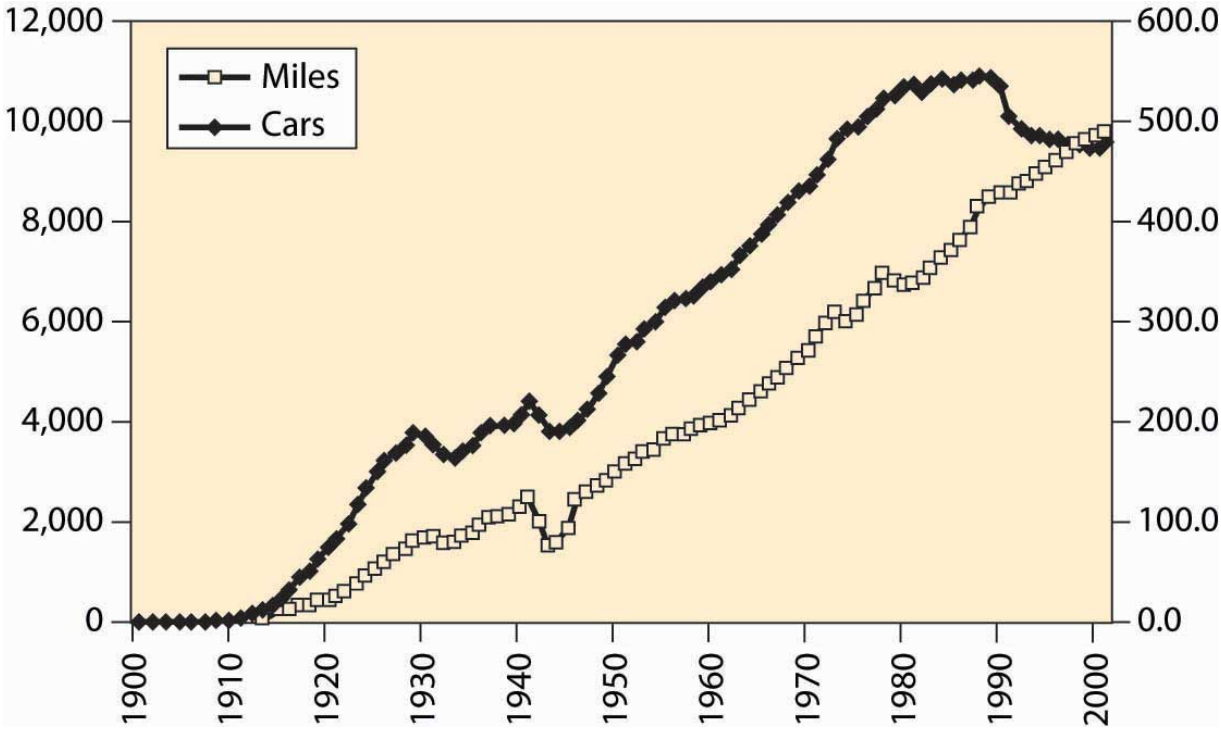


Figure 4.35 Cars per thousand population and miles driven per capita

The cost of selling goods—wholesale and retail costs—remains relatively stable, as does "FIRE," which stands for finance, insurance, and real estate costs. Other services, ranging from restaurants to computer tutoring, have grown substantially. This is the so-called service economy that used to be in the news frequently, but is less so these days.

A bit more than 60% of the population works, with the historical percentage graphed in Figure 4.36. The larger numbers in recent years are partially a reflection of the baby boom's entry into working years, reducing the proportion of elderly and children in American society. However, it is partially a reflection of an increased propensity for households to have two income earners.

Figure 4.36 Percentage of population employed (military and prisoners excluded)

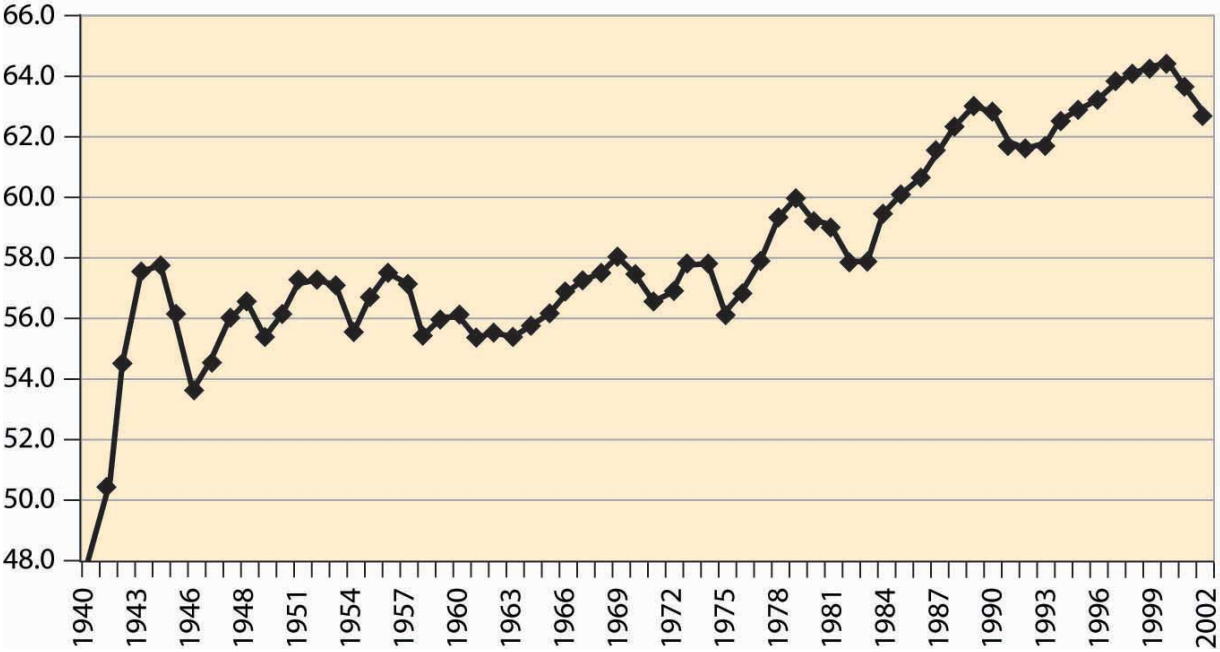


Figure 4.37 Labor force participation rates, all women and married women

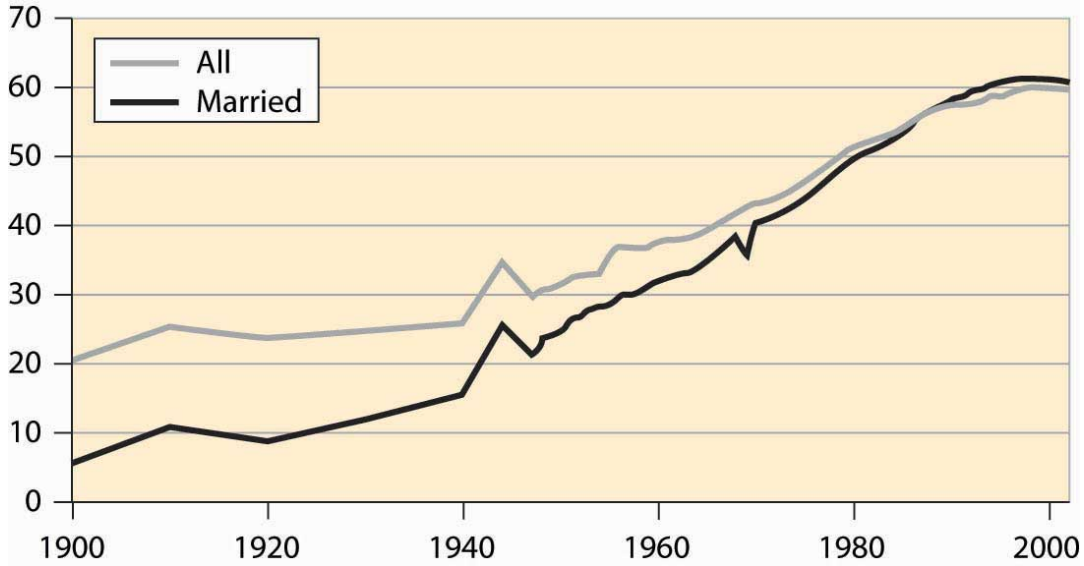


Figure 4.37. The overall participation rate has roughly tripled during the century and significantly exceeds the rate prevailing during World War II, when many women went to work. In addition, participation of married women has now risen above the level for unmarried women. The participation rate for single women is even higher, currently at 68%, and it is higher than the overall average participation rate of all residents. The difference is primarily elderly women, who are disproportionately more likely to be widowed rather than married or single, and who are less likely to be working.

Another sector of the economy that has been of focus in the news is national defense. How much do we spend on the military? In this century, the large expenditure occurred during World War II, when about 50% of GDP was spent by the government and 37% of GDP went to the armed forces. During the Korean War, we spent about 15% of GDP on military goods and less than 10% of GDP during the Vietnam War. The military buildup during Ronald Reagan's presidency (1980–1988) increased our military expenditures from about 5.5% to 6.5% of GDP—a large percentage change in military expenditures, but a small diversion of GDP. The fall of the Soviet Union led the United States to reduce military expenditures, in what was called the "peace dividend," but again the effects were modest, as illustrated in Figure 4.38.

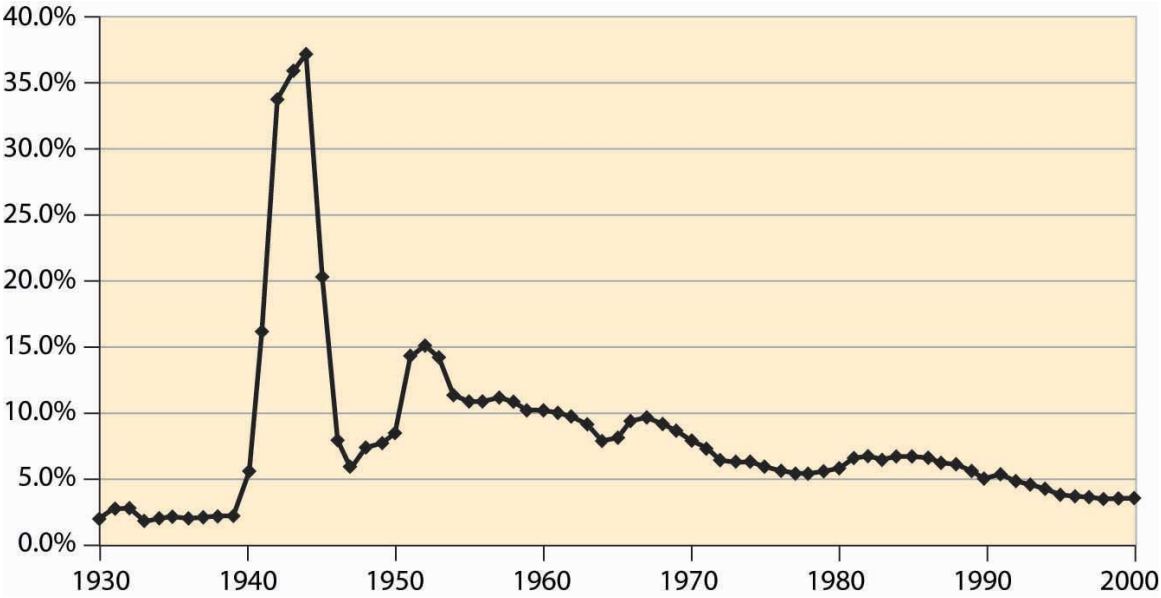


Figure 4.38 Defense as a percentage of GDP

Historically, defense represents the largest expenditure by the federal government. However, as we see, defense has become a much smaller part of the economy overall. Still, the federal government plays many other roles in the modern U.S. economy.

Key Takeaways

- Economists measure output with the gross domestic product (GDP), which is the value of traded goods and services produced within the borders of the United States.
- Importantly, GDP excludes nontraded goods and services. Thus, GDP is not a perfect measure of the production of our society. It is just the best measure we have.
- Economists often use a logarithmic scale rather than a dollar scale. On a logarithmic scale, a straight line gives constant percentage growth.
- Economists divide production into goods and services. Goods are historically divided into mining, construction, and manufacturing.
- The prices of raw materials have fallen even as output and population have grown.
- Manufacturing has become less important for several reasons. Many manufactured goods cost less, pulling down the overall value. In addition, we import more manufactured goods than in the past. We produce more services.
- Electricity production has risen dramatically.
- The number of automobiles per capita in the United States peaked in the early 1980s, but we still drive more than ever, suggesting the change is actually an increase in the reliability of automobiles.
- The cost of selling goods—wholesale and retail costs—remains relatively stable, as does “FIRE” (finance, insurance, and real estate costs). Other services have grown substantially.
- A bit more than 60% of the population works.
- Female participation in the labor force has risen quite dramatically in the United States.
- Military expenditures peaked during World War II, when about 50% of GDP was spent by the government, and 37% of GDP went to the armed forces. During the Korean War, we spent about 15% of GDP on the military and less than 10% of GDP during the Vietnam War. The military buildup during Ronald Reagan’s presidency (1980–1988) increased our military expenditures from about 5.5% to 6.5% of GDP—a large percentage change in military expenditures, but a small diversion of GDP.

This page titled 4.4: Production is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

4.5: Government 4.5: GovernmentEdit Update titlecancel

Learning Objectives

- How big is government and what does the government spend money on?

With a budget over $2 trillion, the federal government represents just under 20% of the U.S. economy. It is one of the largest organizations in the world; only nations are larger organizations, and only a handful of nations are larger.

The size of the federal government, as a percentage of GDP, is shown in Figure 4.39. Federal expenditures boomed during World War II (1940–1945), but shrank back to nearly prewar levels shortly afterward, with much of the difference attributable to veterans' benefits and continuing international involvement. Federal expenditures, as a percentage of GDP, continued to grow until Ronald Reagan's presidency in 1980, when they began to shrink slightly after an initial growth. Figure 4.39 also shows federal revenues, and the deficit—the difference between expenditures and revenues—is apparent, especially for World War II and 1970 to 1998.

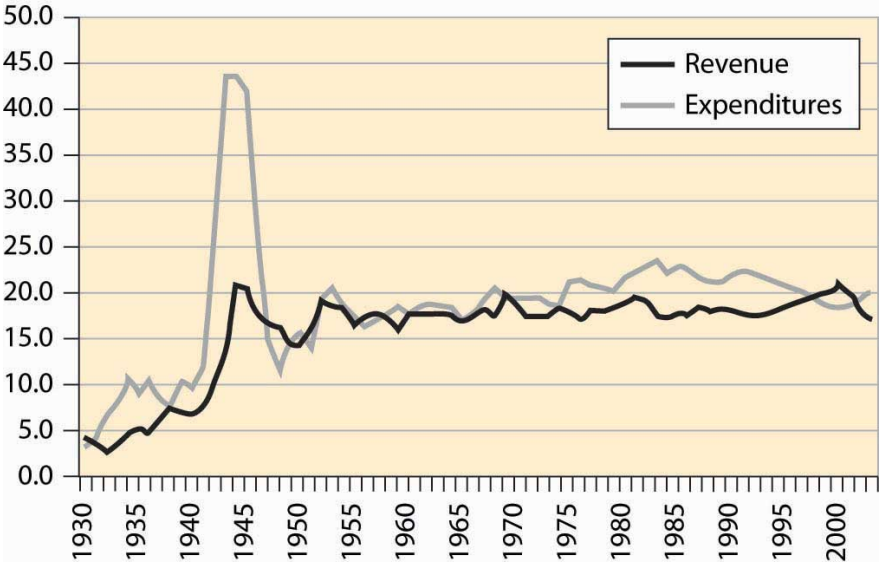


Figure 4.39 Federal expenditures and revenues (percent of GDP)

Much has been written about the federal government's "abdication" of various services, which are pushed onto state and local government. Usually this behavior is attributed to the Reagan presidency (1980–1988). There is some evidence of this behavior in the postwar data, but the effect is very modest and long term. Most of the growth in state and local government occurred between 1947 and 1970, well before the Reagan presidency; state and local government has been stable since then. Moreover, the expenditure of the federal government, which shows ups and downs, has also been fairly stable. In any event, such effects are modest overall.

Figure 4.40 Federal, state, and local level, and total government receipts (% GDP)

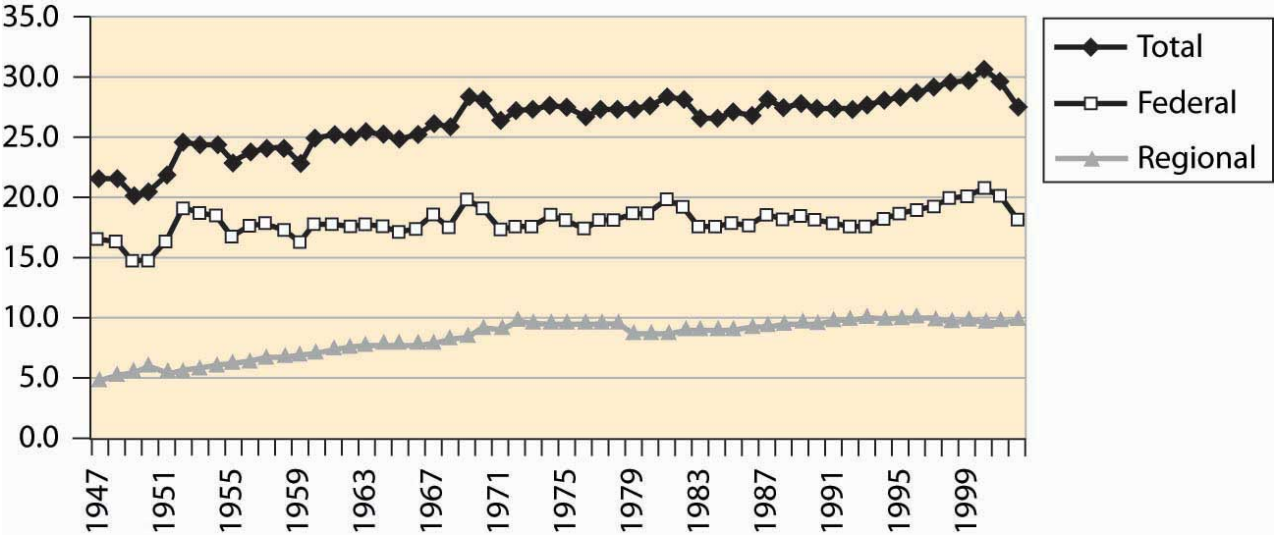


Figure 4.41 shows expenditures of the same entities. Both figures are stated as a percentage of GDP. State and local taxation and expenditures doubled over the postwar period. The two figures are very similar. The federal government's expenditures have varied more significantly than its revenues.

Figure 4.41 Federal, regional, and total expenditures (% GDP)

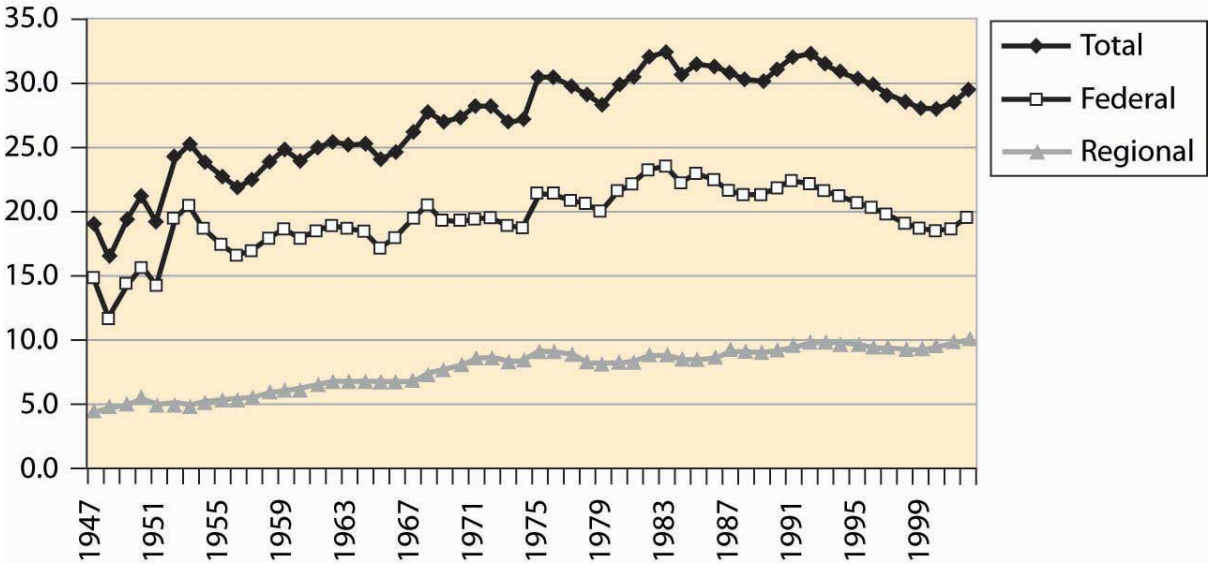


Figure 4.42.

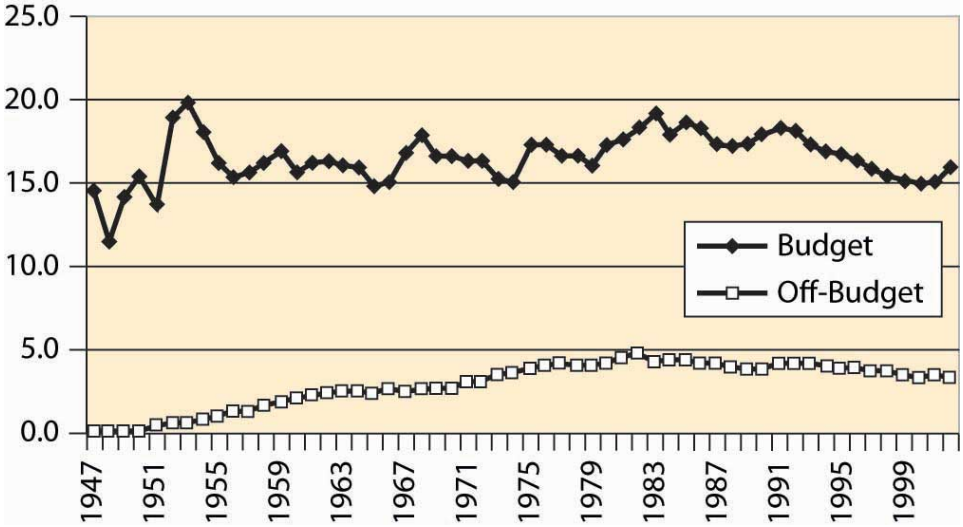


Figure 4.42 Federal expenditures, on and off budget (% GDP)

During the 1980s, the public became aware of off-budget items. Political awareness made off-budget items cease to work as a device for evading balanced-budget requirements, and few new ones were created, although they continue to be debated. Sporadically, there are attempts to push social security off-budget.

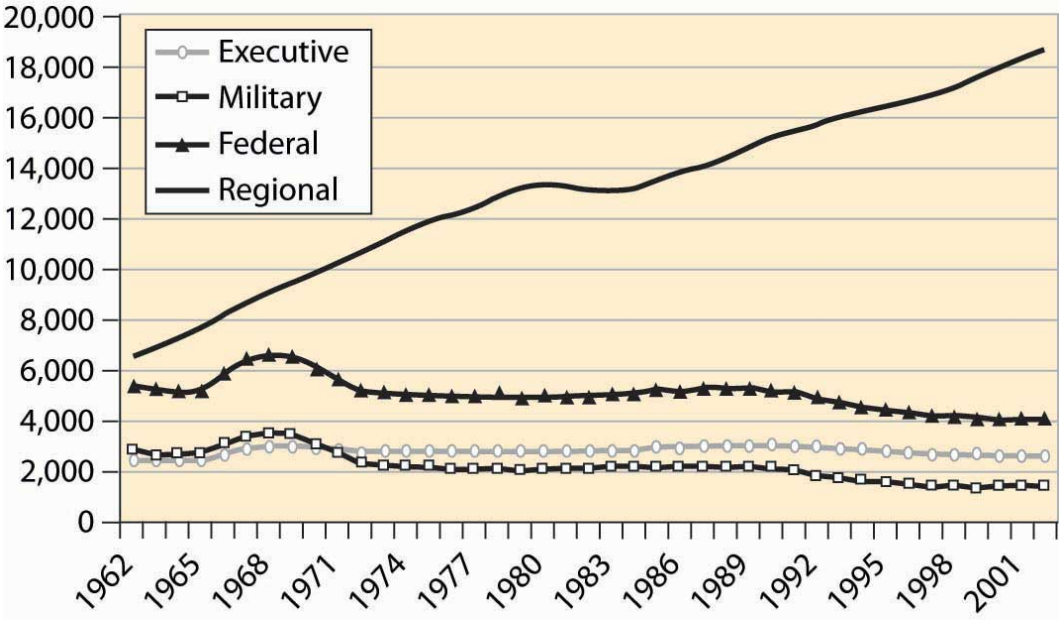


Figure 4.43 Federal and regional government employment

Federal employees include two major categories: uniformed military personnel and the executive branch. State and local government is much larger and has tripled in size since 1962, a fact illustrated in Figure 4.43. The biggest growth areas involve public school teachers, police, corrections (prisons), and hospitals. About 850,000 of the federal employees work for the postal service.

Figure 4.44 Major expenditures of the federal government

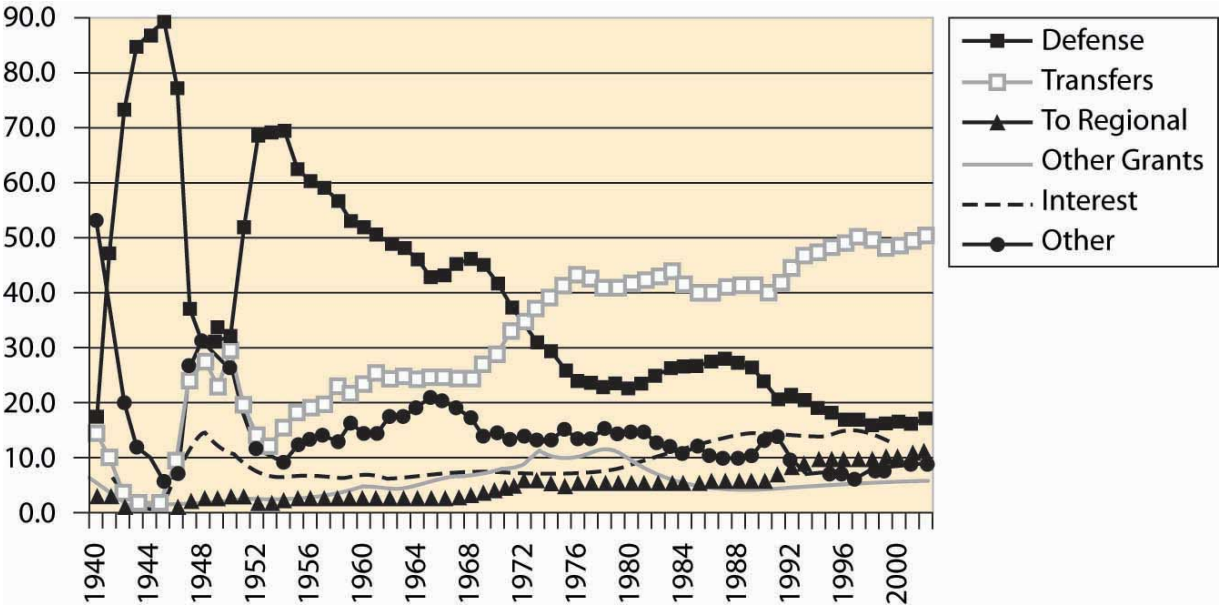


Figure 4.45 Major transfer payments (% of federal budget)

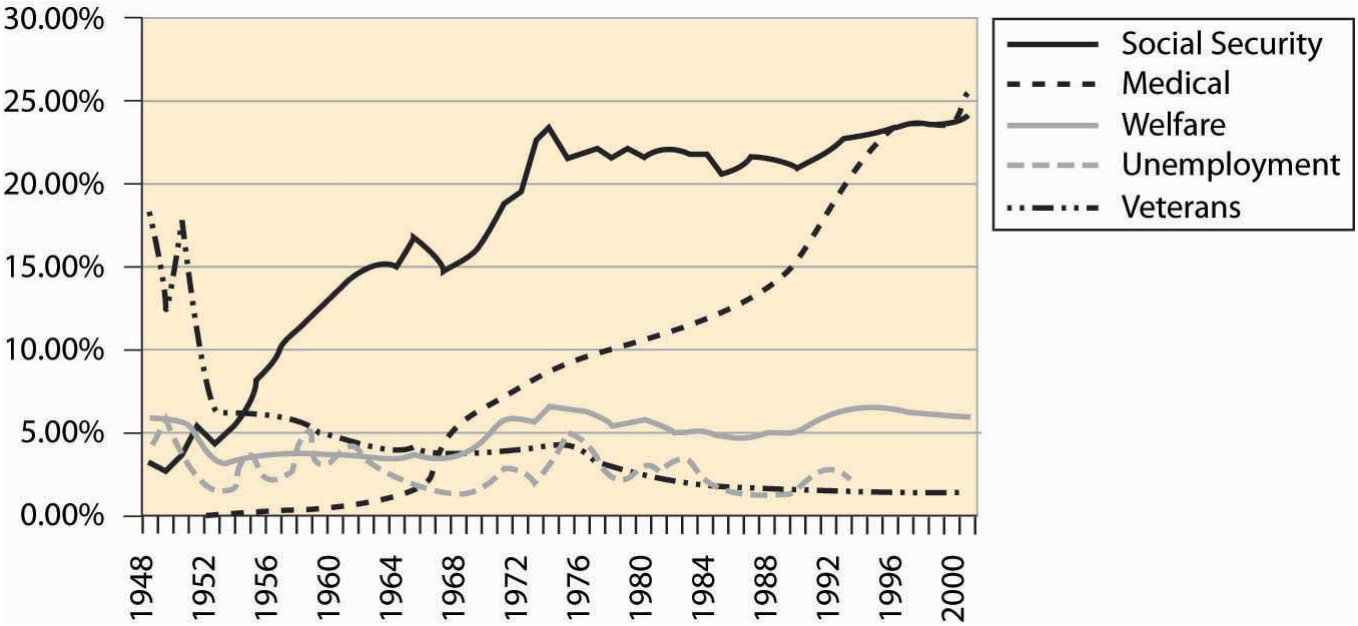


Figure 4.45. The growth in social security during the 1950s and 1960s is primarily a consequence of increasing benefit levels. The growth in Medicare and Medicaid payments over the period 1970 to 1990, in contrast, is primarily a consequence of increased costs of existing programs rather than increases in benefit levels.

Figure 4.46 Social security revenue and expenditure ($ millions)

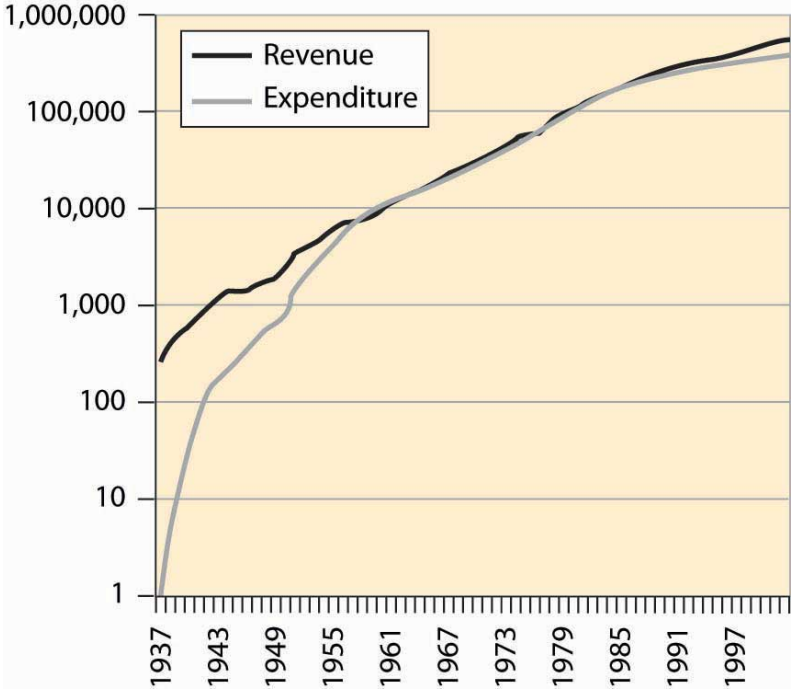


Figure 4.46 makes clear, for most of its existence the social security program has paid out approximately what it took in.

The social security administration has been ostensibly investing money and has a current value of approximately $1.5 trillion, which is a bit less than four times the current annual expenditure on social security. Unfortunately, this money is "invested" in the federal government, and thus is an obligation of the federal government, as opposed to an investment in the stock market. Consequently, from the perspective of someone who is hoping to retire in, say, 2050, this investment isn't much comfort, since the investment won't make it easier for the federal government to make the social security payments. The good news is that the government can print money. The bad news is that when the government prints a lot of money and the obligations of the social security administration are in the tens of trillions of dollars, it isn't worth very much.

Figure 4.47 Federal debt, total and percent of GDP

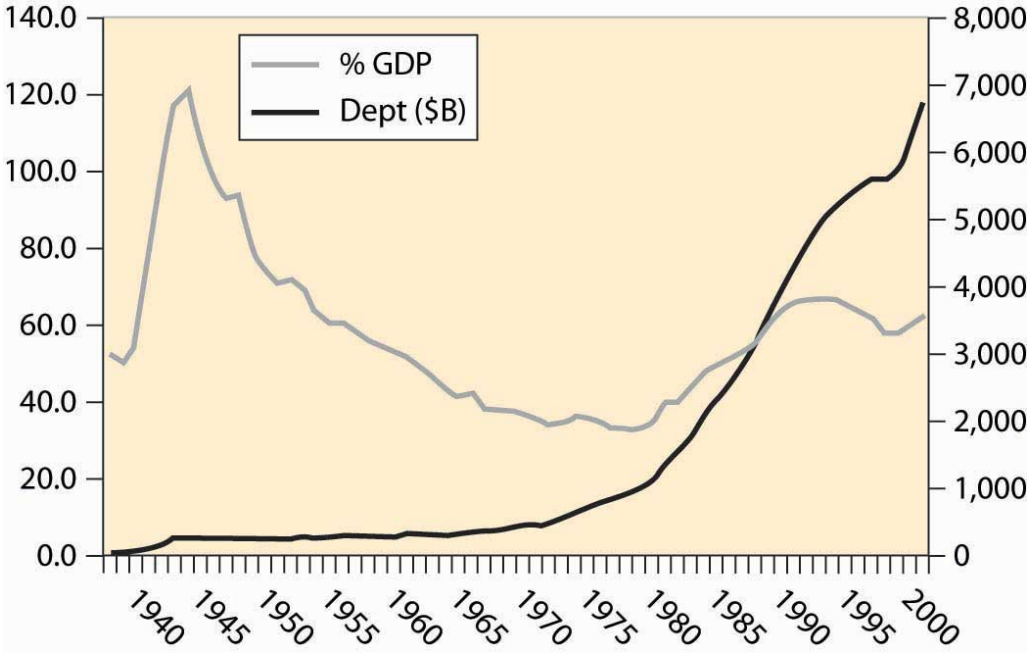


Figure 4.47 gives the size of the federal debt in absolute dollars and as a percent of GDP. The debt was increased dramatically during World War II (1940–1945), but over the next 25 years little was added to it, so that as a portion of growing GDP, the debt fell.

Starting in the late 1970s, the United States began accumulating debt faster than it was growing, and the debt began to rise. That trend wasn't stabilized until the 1990s, and then only because the economy grew at an extraordinary rate by historical standards. The expenditures following the September 11, 2001, terrorist attacks, combined with a recession in the economy, have sent the debt rising dramatically, wiping out the reduction of the 1990s.

The national debt isn't out of control, yet. At 4% interest rates on federal borrowing, we spend about 2.5% of GDP on interest servicing the federal debt. The right evaluation of the debt is as a percentage of GDP; viewed as a percentage, the size of the debt is of moderate size—serious but not critical. The serious side of the debt is the coming retirement of the baby boom generation, which is likely to put additional pressure on the government.

An important distinction in many economic activities is one between a stock and a flow. A **stock** is the current amount of some material; **flow** represents the rate of change in the amount of some material that exists from one instant to the next. Your bank account represents a stock of money; expenditures and income represent a flow. The national debt is a stock; the deficit is the addition to the debt and is a flow. If you think about a lake with incoming water and evaporation, the amount of water in the lake is the stock of water, while the incoming stream minus evaporation is the flow.

Table 4.1 Expenditures on agencies as percent of non-transfer expenditures

Department or Agency	1977	1990	2002
Legislative	0.4	0.4	0.5
Judiciary	0.2	0.3	0.6
Agriculture	2.1	2.2	2.7
Commerce	3.2	0.7	0.7
Education	3.9	3.8	6.7
Energy	3.1	3.2	2.9
Health	3.7	4.6	8.3
Defense	43.8	59.2	46.9
Homeland Security	-	-	4.1
Housing & Urban Dev.	13.4	2.9	4.3
Interior	1.6	1.3	1.4
Justice	1.0	1.7	2.7
Labor	6.1	1.7	1.7
State	0.6	0.9	1.3
Transportation	2.2	2.6	2.1
Treasury	1.7	1.6	1.4
Veterans	2.3	2.6	3.3
Corps of Engineers	1.0	0.6	0.6
Environmental P.A.	1.1	1.1	1.1
Fed Emergency M.A.	0.2	0.4	0.0
GSA	0.2	0.5	0.0
Intl Assistance	2.8	2.7	1.9
NASA	1.6	2.5	2.0
NSF	0.3	0.4	0.7

Table 4.1 gives the expenditures on various agencies, as a percentage of the **discretionary expenditures**, where discretionary is a euphemism for expenditures that aren't transfers. Transfers, which are also known as entitlements, include social security, Medicare, aid to families with dependent children, unemployment insurance, and veteran's benefits. Table 4.1 provides the expenditures by what is sometimes known as the "Alphabet Soup" of federal agencies (DOD, DOJ, DOE, FTC, SEC, …).

The National Science Foundation (NSF) provides funding for basic research. The general idea of government-funded research is that it is useful for ideas to be in the public domain and, moreover, that some research isn't commercially viable but is valuable nevertheless. Studying asteroids and meteors produces little, if any, revenue but could, perhaps, save humanity one day in the event that we needed to deflect a large incoming asteroid. (Many scientists appear pessimistic about actually deflecting an asteroid.) Similarly, research into nuclear weapons might be commercially viable; but, as a society, we don't want firms selling nuclear weapons to the highest bidder. In addition to the NSF, the National Institutes of Health, also a government agency, funds a great deal of research. How much does the government spend on research and development (R&D)? Figure 4.48 shows the history of R&D expenditures. The 1960s "space race" competition between the United States and the Soviet Union led to the greatest federal expenditure on R&D, and it topped 2% of GDP. There was a modest increase during the Reagan presidency (1980–1988) in defense R&D, which promptly returned to earlier levels.

Figure 4.48 Federal spending on R&D (% GDP)

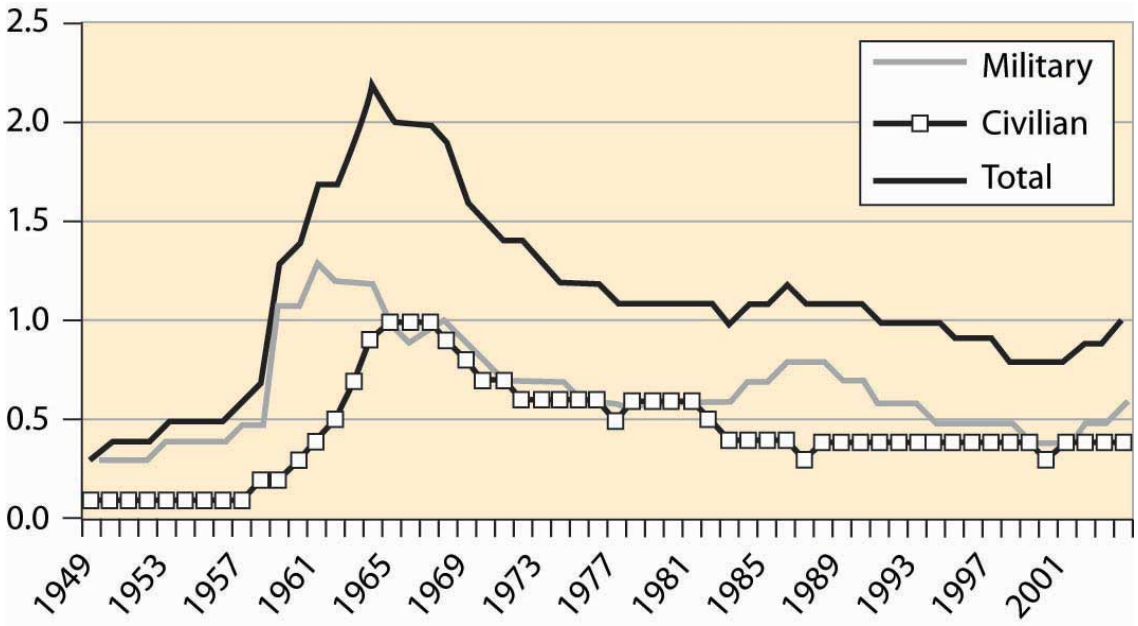


Figure 4.49, the federal income tax currently produces just under 50% of federal revenue. Social security and Medicare taxes produce the next largest portion, with around 30%–35% of revenue. The rest comes from corporate profits' taxes (about 10%), excise taxes like those imposed on liquor and cigarettes (under 5%), and other taxes like tariffs, fees, sales of property like radio spectrum and oil leases, and fines. The major change since World War II is the dramatic increase in social security, a consequence of the federal government's attempt to ensure the future viability of the program, in the face of severely adverse demographics in the form of the retirement of the baby boom generation.

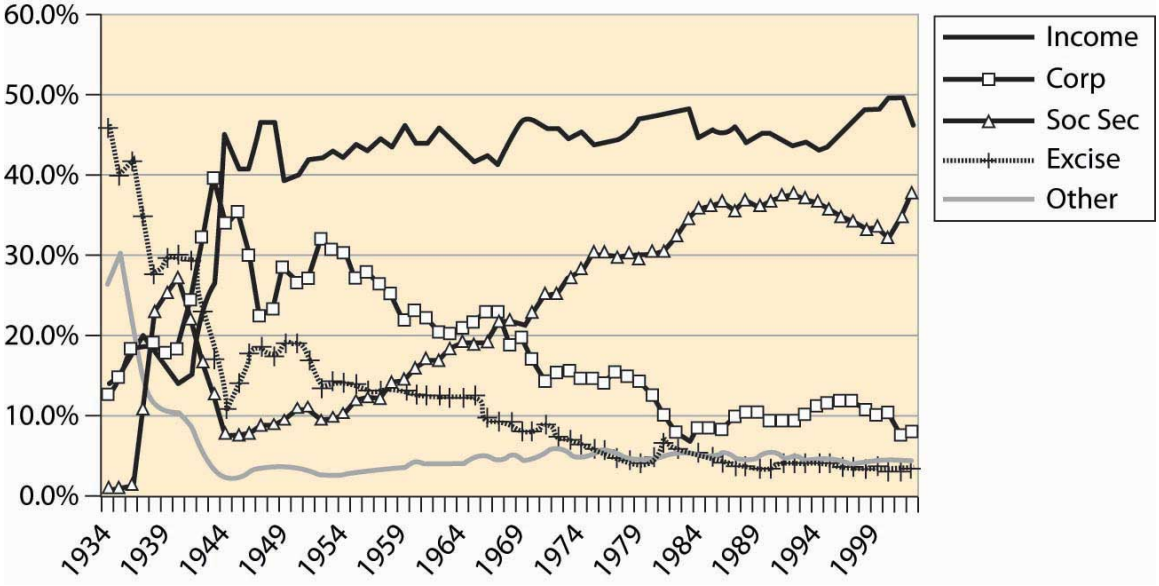


Figure 4.49 Sources of federal government revenue

An important aspect of tax collection is that income taxes, like the federal income tax as well as social security and Medicare taxes, are very inexpensive to collect relative to sales taxes and excise taxes. Income taxes are straightforward to collect even relative to corporate income taxes. Quite reasonably, corporations can deduct expenses and the costs of doing business and are taxed on their profits, not on revenues. What is an allowable deduction, and what is not, makes corporate profits complicated to administer. Moreover, from an economic perspective, corporate taxes are paid by consumers in the form of higher prices for goods, at least when industries are competitive.

Key Takeaways

- With a budget over $2 trillion, the federal government represents just under 20% of the U.S. economy.
- Federal employees include military personnel and the executive branch. State and local government employment is much larger than federal employment and has tripled in size since 1962.
- Transfers to individuals represent almost 50% of federal expenditures. Such transfers include social security, Medicare, Medicaid, unemployment insurance, and veteran's benefits. Transfers are also known as entitlements, and other expenditures are called discretionary spending.
- The social security program has paid out approximately what it took in and is not an investment program.
- The federal government runs deficits, spending more than it earned. Starting in the late 1970s, the United States began accumulating debt faster than it was growing, and the debt began to rise. That trend wasn't stabilized until the 1990s, and then only because the economy grew at an extraordinary rate by historical standards. The expenditures following the September 11, 2001, terrorist attacks, combined with a recession in the economy, have sent the debt rising dramatically.
- The best way to evaluate the debt is as a percentage of GDP.
- An important distinction in many economic activities is one between a stock and a flow. Your bank account represents a stock of money; expenditures and income represent a flow. The national debt is a stock; the deficit is the addition to the debt and is a flow.
- Government funded research and development represents about 1% of GDP, divided about equally between military and civilian research.
- The federal income tax currently produces just under 50% of federal revenue. Social security and Medicare taxes produce the next largest portion, about one third of revenue. The rest comes from corporate profits' taxes (about 10%) and excise taxes like those imposed on liquor and cigarettes (under 5%).

This page titled 4.5: Government is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

4.6: Trade 4.6: TradeEdit Update titlecancel

Learning Objectives

- What do we trade with other nations?

The United States is a major trading nation. Figure 4.50 represents total U.S. imports and exports, including foreign investments and earnings (e.g., earnings from U.S.-owned foreign assets). As is clear from this figure, the net trade surplus ended in the 1970s, and the United States now runs substantial trade deficits, around 4% of the GDP. In addition, trade is increasingly important in the economy.

Figure 4.50 Total imports and exports as a proportion of GDP

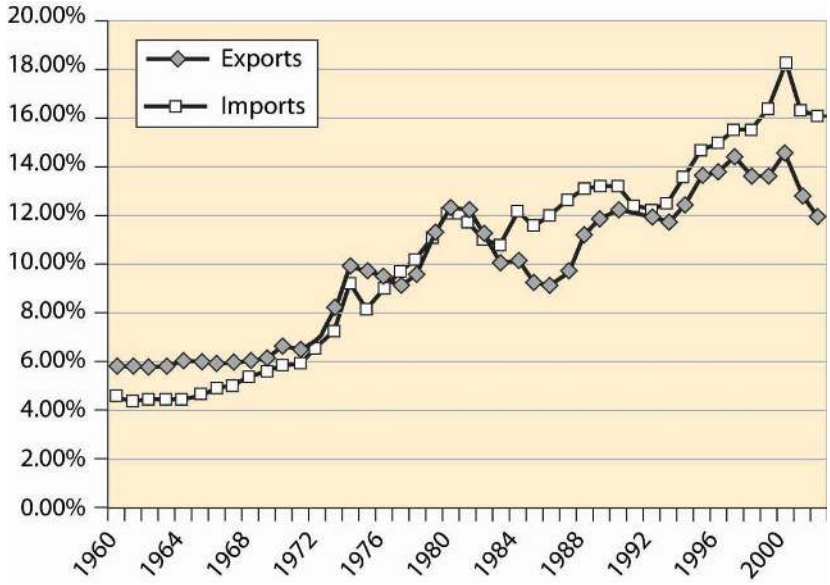


Figure 4.51 shows the total trade in goods and services, as a percentage of U.S. GDP. These figures are surprisingly similar, which show that investments and earnings from investment are roughly balanced—the United States invests abroad to a similar extent as foreigners invest in the United States.

Figure 4.51 U.S. trade in goods and services

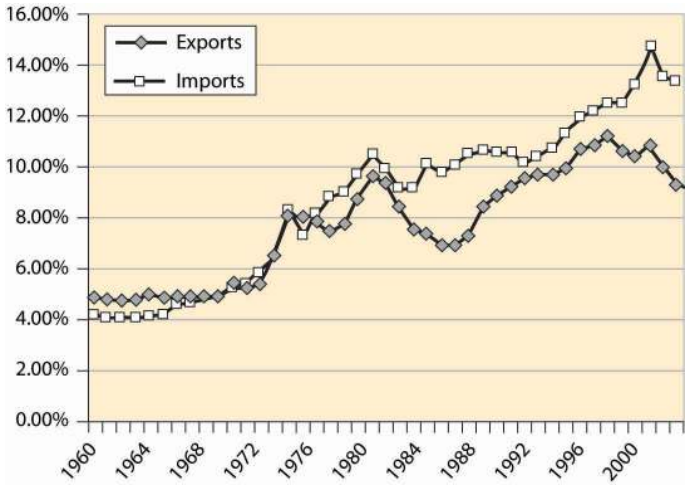


Figure 4.52 shows the earnings on U.S. assets abroad and the payments from U.S.-based assets owned by foreigners. These forms of exchange are known as **capital accounts**. These accounts are roughly in balance, while the United States used to earn about 1% of GDP from its ownership of foreign assets.

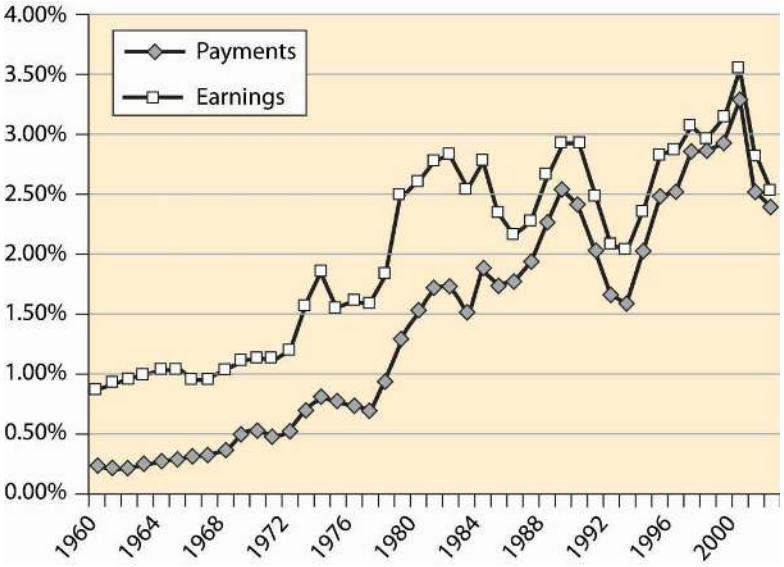


Figure 4.52 Income and payments (% GDP)

Table 4.2 Top U.S. trading partners and trade volumes ($ billions)

Rank	Country	Exports Year-to-Date	Imports Year-to-Date	Total	Percent
	All Countries	533.6	946.6	1,480.2	100.0%
	Top 15 Countries	400.7	715.4	1,116.2	75.4%
1	Canada	123.1	167.8	290.9	19.7%
2	Mexico	71.8	101.3	173.1	11.7%
3	China	22.7	121.5	144.2	9.7%
4	Japan	36.0	85.1	121.0	8.2%
5	Germany	20.4	50.3	70.8	4.8%
6	United Kingdom	23.9	30.3	54.2	3.7%
7	Korea, South	17.5	29.6	47.1	3.2%
8	Taiwan	14.0	22.6	36.5	2.5%
9	France	13.4	20.0	33.4	2.3%
10	Italy	6.9	18.6	25.5	1.7%
11	Malaysia	7.2	18.0	25.2	1.7%
12	Ireland	5.2	19.3	24.5	1.7%
13	Singapore	13.6	10.1	23.7	1.6%
14	Netherlands	15.7	7.9	23.6	1.6%
15	Brazil	9.3	13.2	22.5	1.5%

Who does the United States trade with? Table 4.2 details the top 15 trading partners and the share of trade. The United States and Canada remain the top trading countries of all pairs of countries. Trade with Mexico has grown substantially since the enactment of the 1994 North American Free Trade Act (NAFTA), which extended the earlier U.S.–Canada agreement to include Mexico, and Mexico is the second largest trading partner of the United States. Together, the top 15 account for three quarters of U.S. foreign trade.

Key Takeaways

- The United States is a major trading nation, buying about 16% of GDP and selling about 12%, with a 4% trade deficit. Income from investments abroad is roughly balanced with foreign earnings from U.S. investments; these are known as the capital accounts.
- The United States and Canada remain the top trading countries of all pairs of countries. Mexico is the second largest trading partner of the United States. China and Japan are third and fourth. Together, the top 15 account for three quarters of U.S. international trade.

This page titled 4.6: Trade is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

4.7: Fluctuations 4.7: FluctuationsEdit Update titlecancel

Learning Objectives

- What is a recession?

The U.S. economy has **recessions**, a term that refers to a period marked by a drop in gross domestic output. Recessions are officially called by the National Bureau of Economic Research, which keeps statistics on the economy and engages in various kinds of economic research. Generally, a recession is called whenever output drops for one-half of a year.

Figure 4.53 Postwar industrial production and recessions

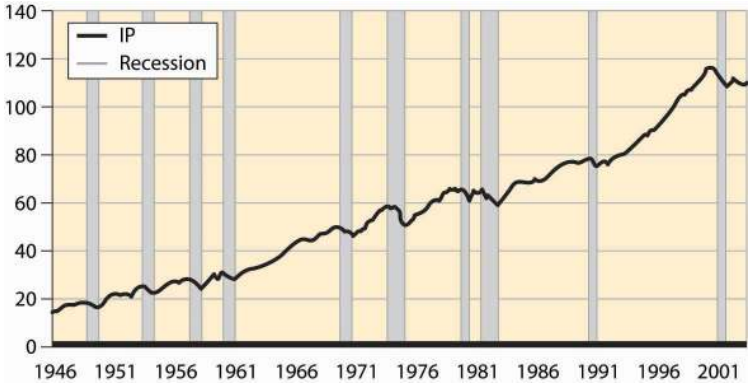


Figure 4.53 shows the overall industrial production of the United States since World War II. Drops in output are clearly noticeable. The official recessions are also marked. There are three booms that lasted about a decade; these are the longest booms in U.S. history and much longer than booms ordinarily last. Prior to World War II, a normal boom lasted 2½ years and the longest boom was 4 years. Recessions have historically lasted for 1½ to 2 years, a pattern that continues. Indeed, the average recession since World War II has been shorter than the average recession prior to that time.

These fluctuations in output are known as the **business cycle**, which is not an exact periodic cycle but instead a random cycle.

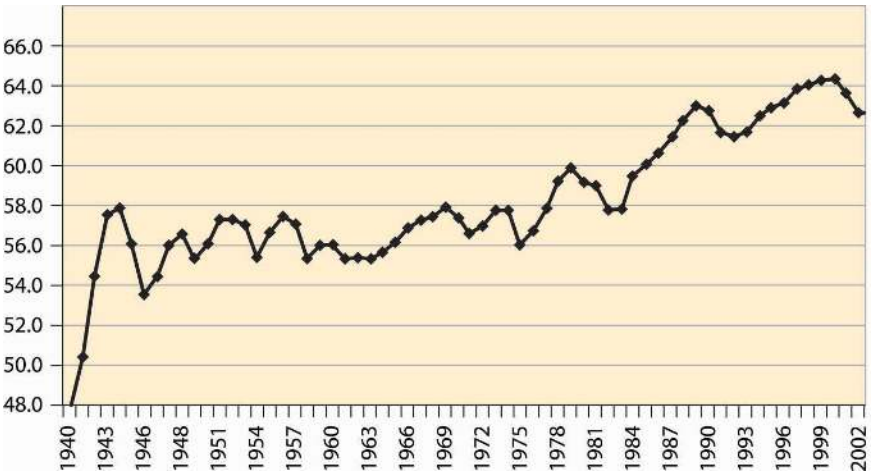


Figure 4.54 Percentage of the population employed

An important aspect of the business cycle is that many economic variables move together, or **covary**. Some economic variables vary less with the business cycle than others. Investment varies very strongly with the business cycle, while overall employment varies weakly. Interest rates, inflation, stock prices, unemployment, and many other variables also vary systematically over the business cycle. Recessions are clearly visible in the percentage of the population employed, as illustrated in Figure 4.54

Some economic variables are much more variable than others. For example, investment, durable goods purchases, and utilization of production capacity vary more dramatically over the business cycle than consumption and employment. Figure 4.55 shows the percentage of industrial capacity utilized to produce manufactured goods. This series is more volatile than production itself and responds more strongly to economic conditions.

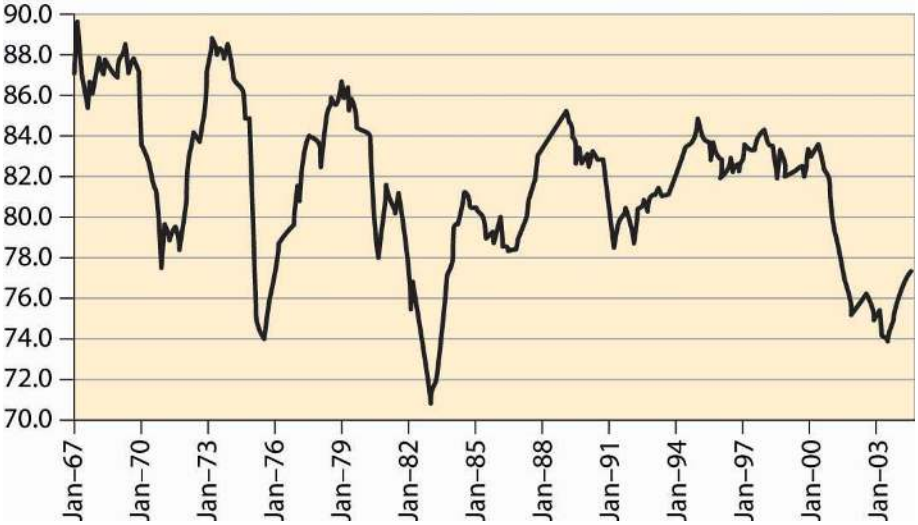


Figure 4.55 Industrial factory capacity utilization

Source: FRED.

Most of the field of macroeconomics is devoted to understanding the determinants of growth and of fluctuations, but further consideration of this important topic is beyond the scope of a microeconomics text.

Key Takeaways

- The U.S. economy has recessions, a term that refers to a drop in gross domestic output. Recessions are officially called by the National Bureau of Economic Research, which keeps statistics on the economy and engages in various kinds of economic research. Generally a recession is called whenever output drops for one half of a year.
- Prior to World War II, a normal boom lasted 2½ years and the longest boom was 4 years, but they have been much longer since 1960. Recessions have historically lasted for 1½ to 2 years, a pattern that continues.
- Fluctuations in output are known as the business cycle, which is not an exact periodic cycle but instead a random cycle.
- An important aspect of the business cycle is that many economic variables move together, or covary. Investment varies very strongly with the business cycle, while overall employment varies weakly. Interest rates, inflation, stock prices, unemployment, and many other variables also vary systematically over the business cycle.
- Some economic variables are much more variable than others. Investment, durable goods purchases, and utilization of production capacity vary more dramatically over the business cycle than consumption and employment.

This page titled 4.7: Fluctuations is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

5: Government Interventions

5.1: Effects of Taxes
5.2: Incidence of Taxes
5.3: Excess Burden of Taxation
5.4: Price Floors and Ceilings
5.5: The Politics of Price Controls
5.6: Price Supports
5.7: Quantity Restrictions and Quotas

This page titled 5: Government Interventions is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

5.1: Effects of Taxes 5.1: Effects of TaxesEdit Update titlecancel

We have so far focused on unimpeded markets, and we saw that markets may perform efficiently.The standard term for an unimpeded market is a *free market*, which is free in the sense of "free of external rules and constraints." In this terminology, eBay is a free market, even though it charges for the use of the market. In this and subsequent chapters, we examine impediments to the efficiency of markets. Some of these impediments are imposed on otherwise efficiently functioning markets, as occurs with taxes. Others, such as monopoly or pollution, impede efficiency in some circumstances, and government may be used to mitigate the problems that arise.

This chapter analyzes taxes. There are a variety of types of taxes, such as income taxes, property taxes, ad valorem (percentage of value) taxes, and excise taxes (taxes on a specific good like cigarettes or gasoline). Here, we are primarily concerned with sales taxes, which are taxes on goods and services sold at retail. Our insights into sales taxes translate naturally into some other taxes.

Learning Objectives

- How do taxes affect equilibrium prices and the gains from trade?

Consider first a fixed, per-unit tax such as a 20-cent tax on gasoline. The tax could either be imposed on the buyer or the supplier. It is imposed on the buyer if the buyer pays a price for the good and then also pays the tax on top of that. Similarly, if the tax is imposed on the seller, the price charged to the buyer includes the tax. In the United States, sales taxes are generally imposed on the buyer—the stated price does not include the tax—while in Canada, the sales tax is generally imposed on the seller.

An important insight of supply and demand theory is that it doesn't matter—to anyone—whether the tax is imposed on the supplier or the buyer. The reason is that ultimately the buyer cares only about the total price paid, which is the amount the supplier gets plus the tax; and the supplier cares only about the net to the supplier, which is the total amount the buyer pays minus the tax. Thus, with a a 20-cent tax, a price of $2.00 to the buyer is a price of $1.80 to the seller. Whether the buyer pays $1.80 to the seller and an additional 20 cents in tax, or pays $2.00, produces the same outcome to both the buyer and the seller. Similarly, from the seller's perspective, whether the seller charges $2.00 and then pays 20 cents to the government, or charges $1.80 and pays no tax, leads to the same profit.There are two minor issues here that won't be considered further. First, the party who collects the tax has a legal responsibility, and it could be that businesses have an easier time complying with taxes than individual consumers. The transaction costs associated with collecting taxes could create a difference arising from who pays the tax. Such differences will be ignored in this book. Second, if the tax is percentage tax, it won't matter to the outcome; but the calculations are more complicated because a 10% tax on the seller at a seller's price of $1.80 is different from a 10% tax on a buyer's price of $2.00. Then the equivalence between taxes imposed on the seller and taxes imposed on the buyer requires different percentages that produce the same effective tax level. In addition, there is a political issue: Imposing the tax on buyers makes the presence and size of taxes more transparent to voters.

First, consider a tax imposed on the seller. At a given price p, and tax t, each seller obtains $p - t$, and thus supplies the amount associated with this net price. Taking the before-tax supply to be S_{Before}, the after-tax supply is shifted up by the amount of the tax. This is the amount that covers the marginal value of the last unit, plus providing for the tax. Another way of saying this is that, at any lower price, the sellers would reduce the number of units offered. The change in supply is illustrated in Figure 5.1.

Figure 5.1 Effect of a tax on supply

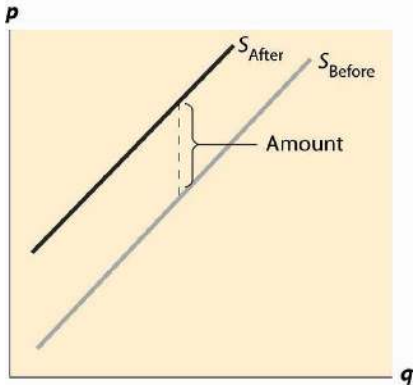


Figure 5.2. In this case, the buyer pays the price of the good, *p*, plus the tax, *t*. This reduces the willingness to pay for any given unit by the amount of the tax, thus shifting down the demand curve by the amount of the tax.

Figure 5.2 Effect of a tax on demand

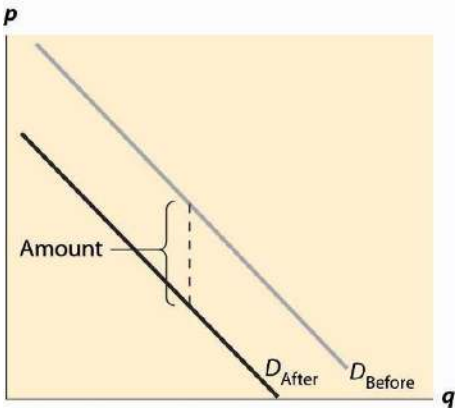


Figure 5.3.

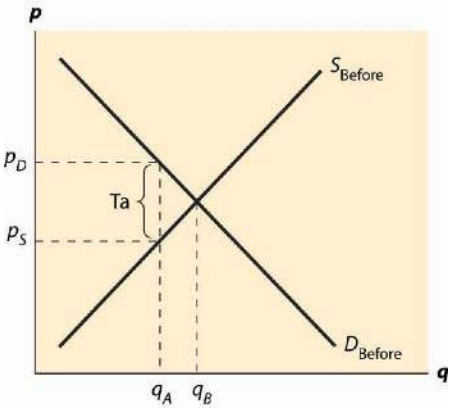


Figure 5.3 Effect of a tax on equilibrium

Also noteworthy in this figure is that the price the buyer pays rises, but generally by less than the tax. Similarly, the price that the seller obtains falls, but by less than the tax. These changes are known as the **incidence of the tax**—a tax mostly borne by buyers, in the form of higher prices, or by sellers, in the form of lower prices net of taxation.

There are two main effects of a tax: a fall in the quantity traded and a diversion of revenue to the government. These are illustrated in Figure 5.4. First, the revenue is just the amount of the tax times the quantity traded, which is the area of the shaded rectangle. The tax raised, of course, uses the after-tax quantity q_A* because this is the quantity traded once the tax is imposed.

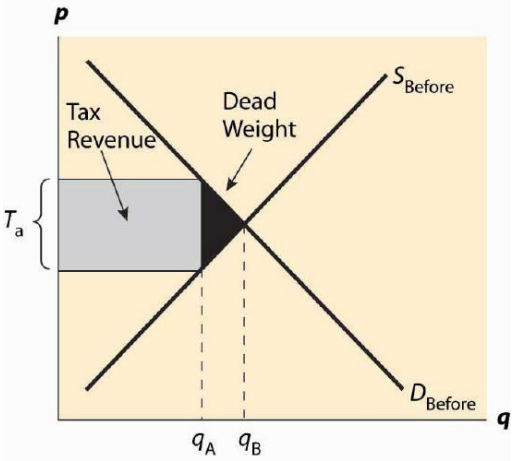


Figure 5.4 Revenue and deadweight loss

In addition, a tax reduces the quantity traded, thereby reducing some of the gains from trade. Consumer surplus falls because the price to the buyer rises, and producer surplus (profit) falls because the price to the seller falls. Some of those losses are captured in the form of the tax, but there is a loss captured by no party—the value of the units that would have been exchanged were there no tax. The value of those units is given by the demand, and the marginal cost of the units is given by the supply. The difference, shaded in black in the figure, is the lost gains from trade of units that aren't traded because of the tax. These lost gains from trade are known as a **deadweight loss**. That is, the deadweight loss is the buyer's values minus the seller's costs of units that are not economic to trade only because of a tax or other interference in the market. The net lost gains from trade (measured in dollars) of these lost units are illustrated by the black triangular region in the figure.

The deadweight loss is important because it represents a loss to society much the same as if resources were simply thrown away or lost. The deadweight loss is value that people don't enjoy, and in this sense can be viewed as an opportunity cost of taxation; that is, to collect taxes, we have to take money away from people, but obtaining a dollar in tax revenue actually costs society more than a dollar. The costs of raising tax revenues include the money raised (which the taxpayers lose), the direct costs of collection, like tax collectors and government agencies to administer tax collection, and the deadweight loss—the lost value created by the incentive effects of taxes, which reduce the gains for trade. The deadweight loss is part of the overhead of collecting taxes. An interesting issue, to be considered in the subsequent section, is the selection of activities and goods to tax in order to minimize the deadweight loss of taxation.

Without more quantification, only a little more can be said about the effect of taxation. First, a small tax raises revenue approximately equal to the tax level times the quantity, or tq. Second, the drop in quantity is also approximately proportional to the size of the tax. Third, this means the size of the deadweight loss is approximately proportional to the tax squared. Thus, small taxes have an almost zero deadweight loss per dollar of revenue raised, and the overhead of taxation, as a percentage of the taxes raised, grows when the tax level is increased. Consequently, the cost of taxation tends to rise in the tax level.

Key Takeaways

- Imposing a tax on the supplier or the buyer has the same effect on prices and quantity.
- The effect of the tax on the supply-demand equilibrium is to shift the quantity toward a point where the before-tax demand minus the before-tax supply is the amount of the tax.
- A tax increases the price a buyer pays by less than the tax. Similarly, the price the seller obtains falls, but by less than the tax. The relative effect on buyers and sellers is known as the incidence of the tax.
- There are two main economic effects of a tax: a fall in the quantity traded and a diversion of revenue to the government.
- A tax causes consumer surplus and producer surplus (profit) to fall.. Some of those losses are captured in the tax, but there is a loss captured by no party—the value of the units that would have been exchanged were there no tax. These lost gains from trade are known as a deadweight loss.
- The deadweight loss is the buyer's values minus the seller's costs of units that are not economic to trade only because of a tax (or other interference in the market efficiency).
- The deadweight loss is important because it represents a loss to society much the same as if resources were simply thrown away or lost.

- Small taxes have an almost zero deadweight loss per dollar of revenue raised, and the overhead of taxation, as a percentage of the taxes raised, grows when the tax level is increased.

EXERCISES

1. Suppose demand is given by $qd(p) = 1 - p$ and supply $qs(p) = p$, with prices in dollars. If sellers pay a 10-cent tax, what is the after-tax supply? Compute the before-tax equilibrium price and quantity, the after-tax equilibrium quantity, and buyer's price and seller's price.
2. Suppose demand is given by $qd(p) = 1 - p$ and supply $qs(p) = p$, with prices in dollars. If buyers pay a 10-cent tax, what is the after-tax demand? Do the same computations as the previous exercise, and show that the outcomes are the same.
3. Suppose demand is given by $qd(p) = 1 - p$ and supply $qs(p) = p$, with prices in dollars. Suppose a tax of t cents is imposed, $t \leq 1$. What is the equilibrium quantity traded as a function of t? What is the revenue raised by the government, and for what level of taxation is it highest?

This page titled 5.1: Effects of Taxes is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

5.2: Incidence of Taxes 5.2: Incidence of TaxesEdit Update titlecancel

Learning Objectives

- Who bears the largest burden of a tax-buyers or sellers?

How much does the quantity fall when a tax is imposed? How much does the buyer's price rise and the price to the seller fall? The elasticities of supply and demand can be used to answer this question. To do so, we consider a percentage tax t and employ the methodology introduced in Chapter 2, assuming constant elasticity of both demand and supply. Let the equilibrium price to the seller be p_s and the equilibrium price to the buyer be p_b. As before, we will denote the demand function by $q_d(p) = ap^{-\varepsilon}$ and supply function by $q_s(p) = bp^{\eta}$. These prices are distinct because of the tax, and the tax determines the difference:

$$p_b = (1+t)p_s$$

Equilibrium requires

$$\text{a } pd - \varepsilon = qd(pb) = qs(ps) = bps\eta$$

Thus,

$$a((1+t)ps) - \varepsilon = apd - \varepsilon = qd(pb) = qs(ps) = bps\eta.$$

This solves for

$$pd = (1+t)ps = (ab)1n + \varepsilon(1+t)\eta\eta + \varepsilon$$

Recall the approximation $(1+t)r \approx 1 + rt$.

Thus, a small proportional tax increases the price to the buyer by approximately η t ε+η and decreases the price to the seller by ε t ε+η . The quantity falls by approximately η ε t ε+η . Thus, the price effect is mostly on the "relatively inelastic party." If demand is inelastic, ε is small; then the price decrease to the seller will be small and the price increase to the buyer will be close to the entire tax. Similarly, if demand is very elastic, ε is very large, and the price increase to the buyer will be small and the price decrease to the seller will be close to the entire tax.

We can rewrite the quantity change as $\eta\varepsilon t\varepsilon + \eta = t1\varepsilon + 1\eta$. Thus, the effect of a tax on quantity is small if either the demand or the supply is inelastic. To minimize the distortion in quantity, it is useful to impose taxes on goods that either have inelastic demand or inelastic supply.

For example, cigarettes are a product with very inelastic demand and moderately elastic supply. Thus, a tax increase will generally increase the price by almost the entire amount of the tax. In contrast, travel tends to have relatively elastic demand, so taxes on travel—airport, hotel, and rental car taxes—tend not to increase the final prices so much but have large quantity distortions.

Key Takeaways

- A small proportional tax t increases the price to the buyer by approximately $\eta t\varepsilon + \eta$ and decreases the price to the seller by $\varepsilon t\varepsilon + \eta$. The quantity falls by approximately $\eta\varepsilon t\varepsilon + \eta$.
- The price effect is mostly on the "relatively inelastic party."
- The effect of a tax on quantity is small if either the demand or the supply is inelastic. To minimize the distortion in quantity, it is useful to impose taxes on goods that either have inelastic demand or inelastic supply.

EXERCISE

1. For the case of constant elasticity (of both supply and demand), what tax rate maximizes the government's revenue? How does the revenue-maximizing tax rate change when demand becomes more inelastic?

This page titled 5.2: Incidence of Taxes is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

5.3: Excess Burden of Taxation 5.3: Excess Burden of TaxationEdit Update titlecancel

Learning Objectives

- How does a tax affect the gains from trade?

The presence of the deadweight loss implies that raising $1 in taxes costs society more than $1. But how much more? This idea—that the cost of taxation exceeds the taxes raised—is known as the **excess burden of taxation**, or just the excess burden. We can quantify the excess burden with a remarkably sharp formula.

To start, we will denote the marginal cost of the quantity q by c(q) and the marginal value by v(q). The elasticities of demand and supply are given by the standard formulae

$$\varepsilon = -\,\mathrm{dq}\; q\; \mathrm{dv}\; v = -v(q)qv'(q)$$

and

$$\eta = \mathrm{dq}\, q\, \mathrm{dc}\, c = c(q)qc'(q)$$

Consider an **ad valorem** (at value) tax that will be denoted by t, meaning a tax on the value, as opposed to a tax on the quantity. If sellers are charging c(q), the ad valorem tax is tc(q), and the quantity q* will satisfy $v\left(q^*\right) = (1+t)c\left(q^*\right)$.

From this equation, we immediately deduce

$$dq^*dt = c\left(q^*\right)v'\left(q^*\right) - (1+t)c'\left(q^*\right) = c\left(q^*\right) - v\left(q^*\right)\varepsilon q^* - (1+t)\left(q^*\right)\eta q^* = -q^*(1+t)(1\varepsilon + 1\eta) = -q^*\varepsilon_{\mathrm{n}}(1+t)(\varepsilon+\eta)$$

Tax revenue is given by Tax = tc(q*)q*.

The effect on taxes collected, Tax, of an increase in the tax rate t is

$$dTaxdt = c(q*)q* + t(c(q*) + q*c'(q*))dq*dt = c(q*)(q* - t(1+1\eta)q*\varepsilon\eta(1+t)(\varepsilon+\eta))$$
$$= c(q*)q*(1+t)(\varepsilon+\eta)((1+t)(\varepsilon+\eta) - t(1+\eta)\varepsilon) = c(q*)q*(1+t)(\varepsilon+\eta)(\varepsilon+\eta-t\eta(\varepsilon-1)).$$

Thus, tax revenue is maximized when the tax rate is t_{max}, given by

$$t_{\max} = \varepsilon + \eta\eta(\varepsilon - 1) = \varepsilon\varepsilon - 1(1\eta + 1\varepsilon)$$

The value ε ε−1 is the monopoly markup rate, which we will meet when we discuss monopoly. Here it is applied to the sum of the inverse elasticities.

The gains from trade (including the tax) is the difference between value and cost for the traded units, and thus is

$$GFT = \int 0g^*v(q) - c(q)dq$$

Thus, the change in the gains from trade as taxes increase is given by

$$dGFTdTax = \partial GFT\partial t\partial Tax\partial t = (v(q*) - c(q*))dq*dtc(q*)q*(1+t)(\varepsilon+\eta)(\varepsilon+\eta-t\eta(\varepsilon-1))$$
$$= -(v(q*) - c(q*))q*\varepsilon\eta(1+t)(\varepsilon+\eta)c(q*)q*(1+t)(\varepsilon+\eta)(\varepsilon+\eta-t\eta(\varepsilon-1)) = -(tc(q*))\varepsilon\eta c(q*)(\varepsilon+\eta-t\eta(\varepsilon-1))$$
$$= -\varepsilon\eta t\varepsilon + \eta - t\eta(\varepsilon-1) = -\varepsilon\varepsilon - 1ttmax - t.$$

The value t_{max} is the value of the tax rate t that maximizes the total tax taken. This remarkable formula permits the quantification of the cost of taxation. The minus sign indicates that it is a loss—the deadweight loss of monopoly, as taxes are raised, and it is composed of two components. First, there is the term ε ε−1 , which arises from the change in revenue as quantity is changed, thus measuring the responsiveness of revenue to a quantity change. The second term provides for the change in the size of the welfare loss triangle. The formula can readily be applied in practice to assess the social cost of taxation, knowing only the tax rate and the elasticities of supply and demand.

The formula for the excess burden is a local formula—it calculates the increase in the deadweight loss associated with raising an extra dollar of tax revenue. All elasticities, including those in t_{max}, are evaluated locally around the quantity associated with the current level of taxation. The calculated value of t_{max} is value given the local elasticities; if elasticities are not constant, this value will not necessarily be the actual value that maximizes the tax revenue. One can think of t_{max} as the projected value. It is sometimes

more useful to express the formula directly in terms of elasticities rather than in terms of the projected value of t_{max}, in order to avoid the potential confusion between the projected (at current elasticities) and actual (at the elasticities relevant to t_{max}) value of t_{max}. This level can be read directly from the derivation shown below:

$$dGFTdTax = -\varepsilon\eta t\varepsilon + \eta - \eta(\varepsilon - 1)t$$

Key Takeaways

- The cost of taxation that exceeds the taxes raised is known as the excess burden of taxation, or just the excess burden.
- Tax revenue is maximized when the tax rate is

$$tmax = \varepsilon\varepsilon - 1(1\eta + 1\varepsilon).$$

- The change in the gains from trade as taxes increase is given by

$$\mathrm{dGFT\ dTax\ } = -\varepsilon\varepsilon - 1\ \mathrm{t\ t\ max\ } - \mathrm{t}$$

.

EXERCISES

1. Suppose both demand and supply are linear, qD= (a – b) p and qS = (c + d) p. A quantity tax is a tax that has a constant value for every unit bought or sold. Determine the new equilibrium supply price pS and demand price pD when a quantity tax of amount t is applied.
2. An ad valorem tax is a proportional tax on value, like a sales tax. Repeat the previous exercise for an ad valorem tax t.
3. Let supply be given by p = q and demand by p = 1 – q. Suppose that a per-unit tax of 0.10 is applied.
    1. What is the change in quantity traded?
    2. Compute the tax revenue and deadweight loss.

This page titled 5.3: Excess Burden of Taxation is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

5.4: Price Floors and Ceilings 5.4: Price Floors and CeilingsEdit Update titlecancel

Learning Objectives

- What happens when the government, not a market, sets the price?

A **price floor** is a minimum price at which a product or service is permitted to sell. Many agricultural goods have price floors imposed by the government. For example, tobacco sold in the United States has historically been subject to a quota and a price floor set by the Secretary of Agriculture. Unions may impose price floors as well. For example, the Screen Actors Guild (SAG) imposes minimum rates for guild members, generally pushing up the price paid for actors above what would prevail in an unconstrained market. (The wages of big-name stars aren't generally affected by SAG because these are individually negotiated.) The most important example of a price floor is the **minimum wage**, which imposes a minimum amount that a worker can be paid per hour.

A **price ceiling** is a maximum price that can be charged for a product or service. Rent control imposes a maximum price on apartments (usually set at the historical price plus an adjustment for inflation) in many U.S. cities. Taxi fares in New York, Washington, DC, and other cities are subject to maximum legal fares. During World War II, and again in the 1970s, the United States imposed price controls to limit inflation, imposing a maximum price for the legal sale of many goods and services. For a long time, most U.S. states limited the legal interest rate that could be charged (these are called **usury laws**), and this is the reason why so many credit card companies are located in South Dakota. South Dakota was the first state to eliminate such laws. In addition, ticket prices for concerts and sporting events are often set below the equilibrium price. Laws prohibiting scalping then impose a price ceiling. Laws preventing scalping are usually remarkably ineffective in practice, of course.

The theory of price floors and ceilings is readily articulated with simple supply and demand analysis. Consider a price floor—a minimum legal price. If the price floor is low enough—below the equilibrium price—there are no effects because the same forces that tend to induce a price equal to the equilibrium price continue to operate. If the price floor is higher than the equilibrium price, there will be a surplus because, at the price floor, more units are supplied than are demanded. This surplus is illustrated in Figure 5.5.

In Figure 5.5, the price floor is illustrated with a horizontal line and is above the equilibrium price. Consequently, at the price floor, a larger quantity is supplied than is demanded, leading to a surplus. There are units that are socially efficient to trade but aren't traded—because their value is less than the price floor. The gains from trade associated with these units, which is lost due to the price floor, represent the deadweight loss.

The price increase created by a price floor will increase the total amount paid by buyers when the demand is inelastic, and otherwise will reduce the amount paid. Thus, if the price floor is imposed in order to be of benefit to sellers, we would not expect to see the price increased to the point where demand becomes elastic, for otherwise the sellers receive less revenue. Thus, for example, if the minimum wage is imposed in order to increase the average wages to low-skilled workers, then we would expect to see the total income of low-skilled workers rise. If, on the other hand, the motivation for the minimum wage is primarily to make low-skilled workers a less effective substitute for union workers, and hence allow union workers to increase their wage demands, then we might observe a minimum wage that is in some sense "too high" to be of benefit to low-skilled workers.

Figure 5.5 A price floor

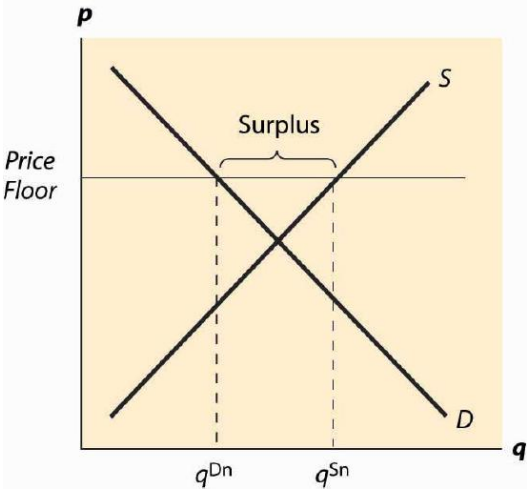


Figure 5.6 Dead weight loss of a price floor

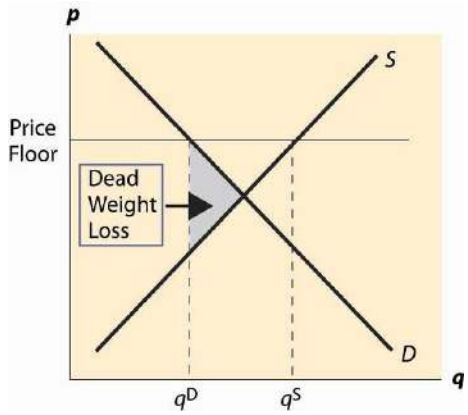


Figure 5.6 is the difference between the value of the units not traded—and value is given by the demand curve—and the cost of producing these units. It is represented by the shaded, triangular-shaped region.

However, this is the minimum loss to society associated with a price floor. Generally there will be other losses. In particular, the loss given above assumes that suppliers who don't sell, don't produce. As a practical matter, some suppliers who won't sell in the end may still produce because they hope to sell. In this case, additional costs are incurred and the deadweight loss will be larger to reflect these costs.

Example: Suppose both supply and demand are linear, with the quantity supplied equal to the price and the quantity demanded equal to one minus the price. In this case, the equilibrium price and the equilibrium quantity are both ½. A price floor of p > ½ induces a quantity demanded of 1 – p. How many units will suppliers offer, if a supplier's chance of trading is random? Suppose $q \geq 1-p$ units are offered. A supplier's chance of selling is 1−p q . Thus, the marginal supplier (who has a marginal cost of q by assumption) has a probability 1−p q of earning p, and a certainty of paying q. Exactly q units will be supplied when this is a break-even proposition for the marginal supplier—that is, $1-pqp-q=0, \text{ or } q=p(1-p)$.

The deadweight loss then includes not just the triangle illustrated in the previous figure, but also the cost of the $p(1-p)-(1-p)$unsold units.

The SAG, a union of actors, has some ability to impose minimum prices (a price floor) for work on regular Hollywood movies. If the SAG would like to maximize the total earnings of actors, what price should they set in the linear demand and supply example?

The effects of a price floor include lost gains from trade because too few units are traded (inefficient exchange), units produced that are never consumed (wasted production), and more costly units produced than necessary (inefficient production).

A price ceiling is a maximum price. Analogous to a low price floor, a price ceiling that is larger than the equilibrium price has no effect. Tell me that I can't charge more than a billion dollars for this book (which is being given away for free), and it won't affect the price charged or the quantity traded. Thus, the important case of a price ceiling is one that is less than the equilibrium price.

In this case, which should now look familiar, the price is forced below the equilibrium price and too few units are supplied, while a larger number are demanded, leading to a shortage. The deadweight loss is illustrated in Figure 5.7, and again represents the loss associated with units that are valued at more than they cost but aren't produced.

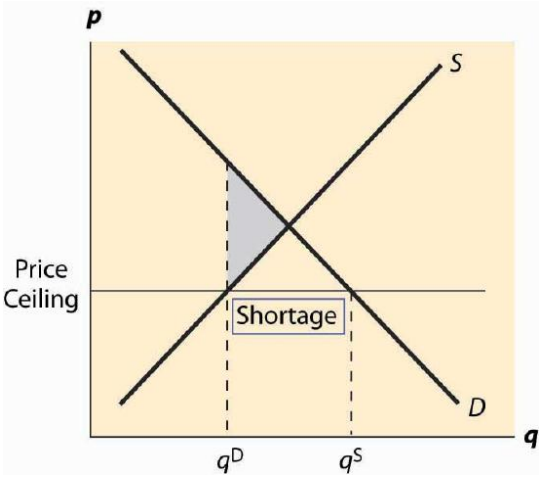


Figure 5.7 A price ceiling

Analogous to the case of a price floor, there can be additional losses associated with a price ceiling. In particular, some lower-value buyers may succeed in purchasing, denying the higher-value buyers the ability to purchase. This effect results in buyers with high values failing to consume, and hence their value is lost.

In addition to the misallocation of resources (too few units and units not allocated to those who value them the most), price ceilings tend to encourage illegal trade as people attempt to exploit the prohibited gains from trade. For example, it became common practice in New York to attempt to bribe landlords to offer rent-controlled apartments, and such bribes could exceed $50,000. In addition, potential tenants expended a great deal of time searching for apartments, and a common strategy was to read the obituaries late at night when the New York Times had just come out, hoping to find an apartment that would be vacant and available for rent.

An important and undesirable by-product of price ceilings is discrimination. In a free or unconstrained market, discrimination against a particular group, based on race, religion, or other factors, requires transacting not based on price but on another factor. Thus, in a free market, discrimination is costly—discrimination entails, for instance, not renting an apartment to the highest bidder but to the highest bidder of the favored group. In contrast, with a price ceiling, there is a shortage; and sellers can discriminate at lower cost, or even at no cost. That is, if there are twice as many people seeking apartments as there are apartments available at the price ceiling, landlords can "pick and choose" among prospective tenants and still get the maximum legal rent. Thus, a price ceiling has the undesirable by-product of reducing the cost of discrimination.

Key Takeaways

- A price floor is a minimum price at which a product or service is permitted to sell. Many agricultural goods have price floors imposed by the government. The most important example of a price floor is the minimum wage.
- A price ceiling is a maximum price that can be charged for a product or service. Rent control imposes a maximum price on apartments in many U.S. cities. A price ceiling that is larger than the equilibrium price has no effect.
- If a price floor is low enough—below the equilibrium price—there are no effects. If the price floor is higher than the equilibrium price, there will be a surplus.
- The deadweight loss of a price floor is the difference between the value of the units not traded—and value is given by the demand curve—and the cost of producing these units. This is the minimum loss to society associated with a price floor.
- The effects of a price floor include lost gains from trade because too few units are traded (inefficient exchange), units produced that are never consumed (wasted production), and more costly units produced than necessary (inefficient production).
- When a price ceiling is below the equilibrium price, the price is forced below the equilibrium price and a shortage results.
- In addition to underproduction, a price ceiling may also lead to inefficient allocation. Price ceilings tend to encourage illegal trade and discrimination.

EXERCISES

1. In Example, show that the quantity produced is less than the equilibrium quantity, which is ½. Compute the gains from trade, given the overproduction of suppliers. What is the deadweight loss of the price floor?
2. Suppose that units aren't produced until after a buyer has agreed to purchase, as typically occurs with services. What is the deadweight loss in this case? (Hint: What potential sellers will offer their services? What is the average cost of supply of this set of potential sellers?)
3. Adapt the price floor example above to the case of a price ceiling, with $p < ½$, and compute the lost gains from trade if buyers willing to purchase are all able to purchase with probability qS/qD. (Hint: Compute the value of qD units; the value realized by buyers collectively will be that amount times the probability of trade.)

This page titled 5.4: Price Floors and Ceilings is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

5.5: The Politics of Price Controls 5.5: The Politics of Price ControlsEdit Update titlecancel

Learning Objectives

- Why does the government meddle in markets?

Both demand and supply tend to be more elastic in the long run. This means that the quantity effects of price floors and ceilings tend to be larger over time. An extreme example of this is rent control, a maximum price imposed on apartments.

Rent control is usually imposed in the following way: As a prohibition or limitation on price increases. For example, New York City's rent control, imposed during World War II, prevented landlords from increasing rent, even when their own costs increased, such as when property taxes increased. This law was softened in 1969 to be gradually replaced by a rent-stabilization law that permitted modest rent increases for existing tenants.

Figure 5.8 Rent control, initial effect

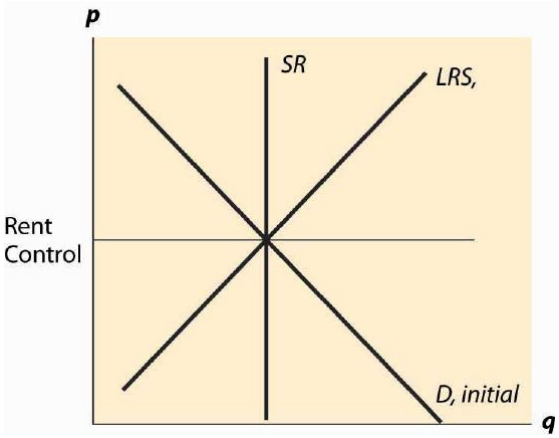


Figure 5.8.

We start with a rent-control law that has little or no immediate effect because it is set at current rents. Thus, in the near term, tenants' fears of price increases are eased and there is little change in the apartment rental market. This is not to say that there is zero effect—some companies considering construction of an apartment building on the basis of an expectation of higher future rents may be deterred, and a few marginal apartments may be converted to other uses because the upside potential for the owner has been removed, but such effects are modest at best.

Figure 5.9 Rent control, long-run effect

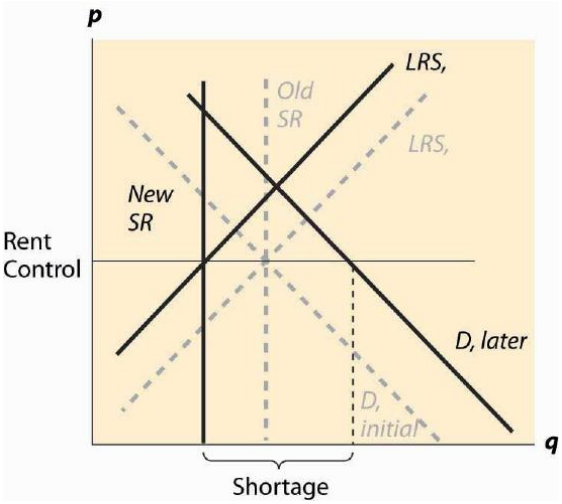


Figure 5.9. Dashed gray lines illustrate the old values of demand, short-run supply, and long-run supply. The new values, reflecting an increase in demand, a fall in long-run supply, and a reduction in the number of available apartments (where the rent control covers the long-run cost) are shown in dark black lines.

The shortage is created by two separate factors—demand is increasing as incomes and population rise, and supply is decreasing as costs rise. This reduces the quantity of available housing units supplied and increases the demand for those units.

How serious is the threat that units will be withdrawn from the market? In New York City, over 200,000 apartment units were abandoned by their owners, usually because the legal rent didn't cover the property taxes and legally mandated maintenance. In some cases, tenants continued to inhabit the buildings even after the electricity and water were shut off. It is fair to say that rent control devastated large areas of New York City, such as the Bronx. So why would New York City, and so many other communities, impose rent control on itself?

The politics of rent control are straightforward. First, rent control involves a money transfer from landlords to tenants, because tenants pay less than they would absent the law, and landlords obtain less revenue. In the short run, due to the inelastic short-run supply, the effect on the quantity of apartments is small, so rent control is primarily just a transfer from landlords to tenants.

In a city like New York, the majority of people rent. A tiny fraction of New Yorkers are landlords. Thus, it is easy to attract voters to support candidates who favor rent control—most renters will benefit, while landlords don't. The numbers, of course, don't tell the whole story because, while landlords are small in number, they are wealthier on average, and thus likely have political influence beyond the number of votes they cast. However, even with their larger economic influence, the political balance favors renters. In the 100*ab* zip codes of Manhattan (the first three digits are 100), 80% of families were renters in the year 2000. Thus, a candidate who runs on a rent-control platform appeals to a large portion of the voters.

Part of the attraction of rent control is that there is little economic harm in the short run, and most of that harm falls on new residents of New York City. As new residents generally haven't yet voted in New York, potential harm to them has only a small effect on most existing New Yorkers, and thus isn't a major impediment to getting voter support for rent control. The slow rate of harm to the city is important politically because the election cycle encourages a short time horizon—if successful at lower office, a politician hopes to move on to higher office and is unlikely to be blamed for the long-run damage to New York City by rent control.

Rent control is an example of a political situation sometimes called the **tyranny of the majority**, where a majority of the people have an incentive to confiscate the wealth of a minority. But there is another kind of political situation that is in some sense the reverse, where a small number of people care a great deal about something, and the majority are only slightly harmed on an individual basis. No political situation appears more extreme in this regard than that of refined sugar. There are few U.S. cane sugar producers (nine in 1997), yet the U.S. imposes quotas that raise domestic prices much higher than world prices, in some years tripling the price that Americans pay for refined sugar. The domestic sugar producers benefit, while consumers are harmed. But consumers are harmed by only a small amount each—perhaps 12 to 15 cents per pound—which is not enough to build a consensus to defeat politicians who accept donations from sugar producers. This is a case where **concentrated benefits and diffused costs** determine the political outcome. A small number of people with strong incentives are able to expropriate a small amount per person from a large number of people. Because there aren't many sugar producers, it is straightforward for them to act as a single force. In contrast, it is pretty hard for consumers to become passionate about 12 cents per pound increase in the domestic sugar price when they consume about 60 pounds per year of sugar.

Key Takeaways

- Both demand and supply tend to be more elastic in the long run.
- Rent control is usually imposed as a prohibition or limitation on price increases. The nature of rent control is that it begins with, at most, minor effects because it doesn't bind until the equilibrium rent increases. Thus, the cost of rent control tends to be in the future, and ill effects worsen over time.
- A candidate who runs on a rent-control platform appeals to a large portion of the voters as there are more renters than landlords.
- Rent control is an example of a political situation sometimes called the tyranny of the majority, where a majority of people have an incentive to confiscate the wealth of a minority.
- Concentrated benefits and diffused costs are the opposite of tyranny of the majority.

This page titled 5.5: The Politics of Price Controls is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

5.6: Price Supports 5.6: Price SupportsEdit Update titlecancel

Learning Objectives

- How is a price support different from a price floor?

A **price support** is a combination of two programs—a minimum price, or price floor, and government purchase of any surplus. Thus, a price support is different from a price floor because, with a price floor, any excess production by sellers is a burden on the sellers. In contrast, with a price support, any excess production is a burden on the government.

The U.S. Department of Agriculture operates a price support for cheese and has possessed warehouses full of cheese in the past. There are also price supports for milk and other agricultural products.

Figure 5.10 Price supports

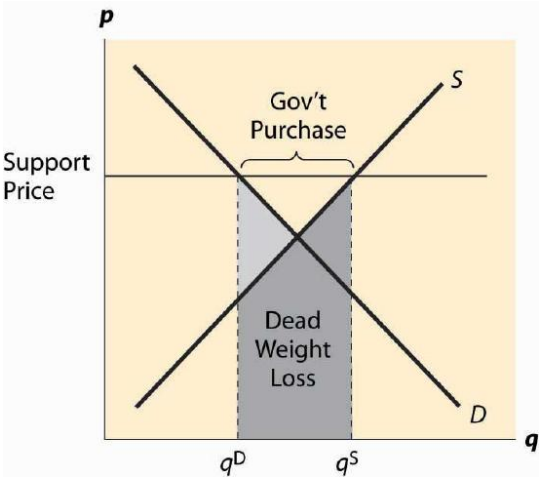


Figure 5.10 illustrates the effect of a support program. The government posts a minimum price it is willing to pay for a product, called the **support price**, and purchases any excess production offered on the market. The government purchases, which are the difference between the quantity supplied and quantity demanded, are illustrated in the figure. The cost of the program to the government is the support price times the quantity purchased, which is the area of the rectangle directly underneath the words "Gov't Purchases."

There are two kinds of deadweight loss in a price-support program. First, consumers who would like to buy at the equilibrium price are deterred by the higher prices, resulting in the usual deadweight loss, illustrated by the lighter shading. In addition, however, there are goods produced that are then either destroyed or put in warehouses and not consumed, which means the costs of production of those goods is also lost, resulting in a second deadweight loss. That loss is the cost of production, which is given by the supply curve, and is the area under the supply curve for the government purchases. It is shaded in a horizontal fashion. The total deadweight loss of the price support is the sum of these two individual losses. Unlike the case of a price floor or ceiling, a price support creates no ambiguity about what units are produced, or which consumers are willing and able to buy. Thus, the rationing aspect of a price floor or ceiling is not present for a price support, nor is the incentive to create a black market other than one created by selling the warehouse full of product.

Key Takeaways

- A price support is a combination of two programs: a price floor and government purchase of surplus. Excess production is a burden on the government.
- A price support above the equilibrium price leads to a surplus.
- The deadweight loss of price supports involves the usual deadweight loss plus the entire cost of unconsumed goods.

This page titled 5.6: Price Supports is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

5.7: Quantity Restrictions and Quotas 5.7: Quantity Restrictions and QuotasEdit Update titlecancel

Learning Objectives

- What is a quota?

The final common way that governments intervene in market transactions is to impose a quota. A **quota** is a maximal production quantity, usually set based on historical production. In tobacco, peanuts, hops, California oranges, and other products, producers have production quotas based on their historical production. Tobacco quotas were established in the 1930s, and today a tobacco farmer's quota is a percentage of the 1930s level of production. The Secretary of Agriculture sets the percentage annually. Agricultural products are not the only products with quotas. The right to drive a taxi in New York requires a medallion issued by the city, and there are a limited number of medallions. This is a quota. Is it a restrictive quota? The current price of a New York taxi medallion—the right to drive a taxi legally in New York City—is $413,000 (as of 2008). This adds approximately $30,000 to $40,000 annually to the cost of operating a taxi in New York, using a risk-adjusted interest rate.

What are the effects of a quota? A quota restricts the quantity below what would otherwise prevail, forcing the price up, which is illustrated in Figure 5.11. It works like a combination of a price floor and a prohibition on entry.

Generally, the immediate effects of a quota involve a transfer of money from buyers to sellers. The inefficient production and surplus of the price floor are avoided because a production limitation created the price increase. This transfer has an undesirable and somewhat insidious attribute. Because the right to produce is a capital good, it maintains a value, which must be captured by the producer. For example, an individual who buys a taxi medallion today, and pays $400,000, makes no economic profits—he captures the forgone interest on the medallion through higher prices but no more than that. The individuals who receive the windfall gain are those who were driving taxis and were grandfathered in to the system and issued free medallions. Those people who were driving taxis 70 years ago—and are mostly dead at this point—received a windfall gain from the establishment of the system. Future generations pay for the program, which provides no net benefits to the current generation. All the benefits were captured by people long since retired.

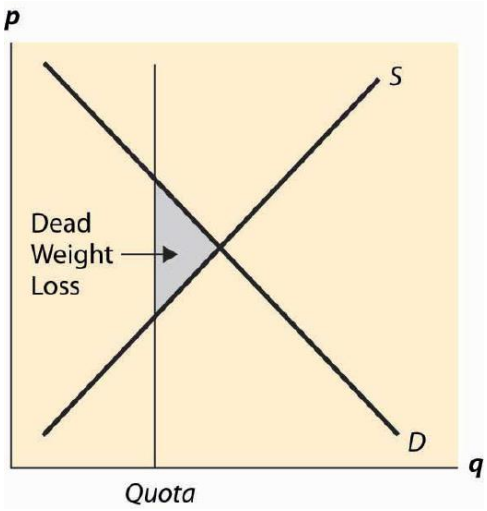


Figure 5.11 A quota

Does this mean that it is harmless to eliminate the medallion requirement? Unfortunately, not. The current medallion owners who, if they bought recently, paid a lot of money for their medallions would see the value of these investments destroyed. Thus, elimination of the program would harm current medallion owners.

If the right to produce is freely tradable, the producers will remain the efficient producers, and the taxi medallions are an example of this. Taxi medallions can be bought and sold. Moreover, a medallion confers the right to operate a taxi, but doesn't require that the owner of the medallion actually drive the taxi. Thus, a "medallion owning company" can lease the right to drive a taxi to an efficient driver, thereby eliminating any inefficiency associated with the person who drives the taxi.

In contrast, because tobacco-farming rights aren't legally tradable across county lines, tobacco is very inefficiently grown. The average size of a burley tobacco farm is less than 5 acres, so some are much smaller. There are tobacco farms in Florida and Missouri, which only exist because of the value of the quota—if they could trade their quota to a farm in North Carolina or Kentucky, which are much better suited to producing cigarette tobacco, it would pay to do so. In this case, the quota, which locked in production rights, also locked in production that gets progressively more inefficient as the years pass.

Quotas based on historical production have the problem that they don't evolve in ways that production methods and technology do, thus tending to become progressively more inefficient. Tradable quotas eliminate this particular problem but continue to have the problem that future generations are harmed with no benefits.

Key Takeaways

- A quota is a maximum production quantity, usually set based on historical production.
- A quota restricts the quantity below what would otherwise prevail, forcing the price up.
- A quota transfers wealth from buyers to sellers. No surplus arises because of the production limitation. Future generations pay for the program, which provides future sellers no benefits.
- Quotas based on historical production have the problem that they don't evolve in ways that production methods and technology do, thus tending to become progressively more inefficient. Tradable quotas eliminate this particular problem, but continue to have the problem that future generations are harmed with no benefits.

EXERCISE

1. Suppose demand for a product is $q_d = 1 - p$, and the marginal cost of production is c. A quota at level $Q \leq 1 - c$ is imposed. What is the value of the quota, per unit of production? Use this to derive the demand for the quota as a function of the level of quota released to the market. If the government wishes to sell the quota, how much should it sell to maximize the revenue on the product?

This page titled 5.7: Quantity Restrictions and Quotas is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

6: Trade

6.1: Production Possibilities Frontier
6.2: Comparative and Absolute Advantage
6.3: Factors of Production
6.4: International Trade

This page titled 6: Trade is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

6.1: Production Possibilities Frontier 6.1: Production Possibilities FrontierEdit Update titlecancel

Supply and demand offers one approach to understanding trade, and it represents the most important and powerful concept in the toolbox of economists. However, for some issues, especially those of international trade, another related tool is very useful: the production possibilities frontier. Analysis using the production possibilities frontier was made famous by the “guns and butter” discussions of World War II. From an economic perspective, there is a trade-off between guns and butter—if a society wants more guns, it must give up something, and one thing to give up is butter. While the notion of getting more guns might lead to less butter often seems mysterious, butter is, after all, made with cows, and indirectly with land and hay. But the manufacture of butter also involves steel containers, tractors to turn the soil, transportation equipment, and labor, all of which either can be directly used (steel, labor) or require inputs that could be used (tractors, transportation) to manufacture guns. From a production standpoint, more guns entail less butter (or other things).

Learning Objectives

- What can we produce, and how does that relate to cost?

Formally, the set of production possibilities is the collection of “feasible outputs” of an individual, group or society, or country. You could spend your time cleaning your apartment, or you could study. The more time you devote to studying, the higher your grades will be, but the dirtier your apartment will be. This is illustrated, for a hypothetical student, in Figure 6.1.

The production possibilities set embodies the feasible alternatives. If you spend all your time studying, you could obtain a 4.0 (perfect) grade point average (GPA). Spending an hour cleaning reduces the GPA, but not by much; the second hour reduces it by a bit more, and so on.

The boundary of the production possibilities set is known as the production possibilities frontier. This is the most important part of the production possibilities set because, at any point strictly inside the production possibilities set, it is possible to have more of everything, and usually we would choose to have more.To be clear, we are considering an example with two goods: cleanliness and GPA. Generally there are lots of activities, like sleeping, eating, teeth brushing, and so on; the production possibilities frontier encompasses all of these goods. Spending all your time sleeping, studying, and cleaning would still represent a point on a three-dimensional frontier. The slope of the production possibilities frontier reflects opportunity cost because it describes what must be given up in order to acquire more of a good. Thus, to get a cleaner apartment, more time or capital, or both, must be spent on cleaning, which reduces the amount of other goods and services that can be had. For the two-good case in Figure 6.1, diverting time to cleaning reduces studying, which lowers the GPA. The slope dictates how much lost GPA there is for each unit of cleaning.

Figure 6.1 The production possibilities frontier

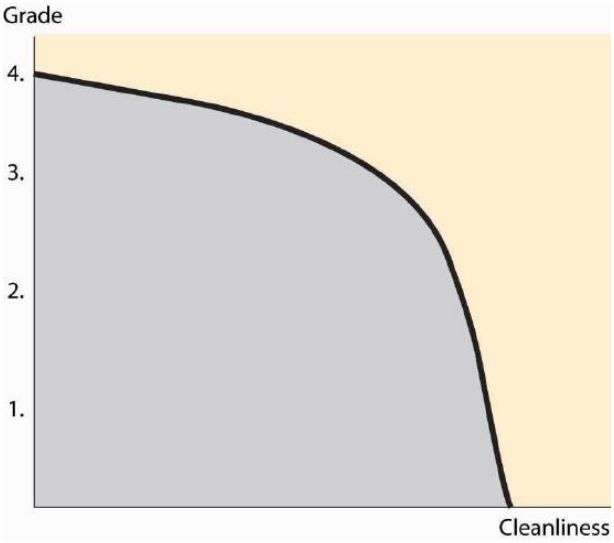


Figure 6.1: They are concave toward the origin. While this feature need not be universally true, it is a common feature, and there is a reason for it that we can see in the application. If you are only going to spend an hour studying, you spend that hour doing the most important studying that can be done in an hour, and thus get a substantial improvement in grades for the hour's work. The second hour of studying produces less value than the first, and the third hour less than the second. Thus, spending more on something reduces the per-unit value produced. This is the principle of Diminishing marginal returns. Diminishing marginal returns are like picking apples. If you are only going to pick apples for a few minutes, you don't need a ladder because the fruit is low on the tree; the more time spent, the fewer apples per hour you will pick.

Consider two people, Ann and Bob, getting ready for a party. One is cutting up vegetables; the other is making hors d'oeuvres. Ann can cut up 2 ounces of vegetables per minute, or make one hors d'oeuvre in a minute. Bob, somewhat inept with a knife, can cut up 1 ounce of vegetables per minute, or make 2 hors d'oeuvres per minute. Ann and Bob's production possibilities frontiers are illustrated in Figure 6.2, given that they have an hour to work.

Since Ann can produce 2 ounces of chopped vegetables in a minute, if she spends her entire hour on vegetables, she can produce 120 ounces. Similarly, if she devotes all her time to hors d'oeuvres, she produces 60 of them. The constant translation between the two means that her production possibilities frontier is a straight line, which is illustrated on the left side of Figure 6.2. Bob's is the reverse—he produces 60 ounces of vegetables or 120 hors d'oeuvres, or something on the line in between.

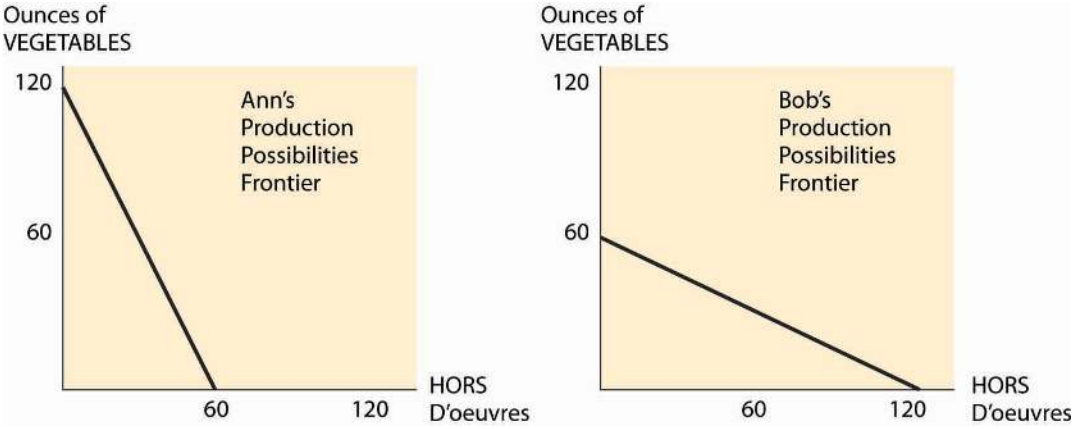


Figure 6.2 Two production possibilities frontiers

For Ann, the opportunity cost of an ounce of vegetables is half of one hors d'oeuvre—to get one extra ounce of vegetables, she must spend 30 extra seconds on vegetables. Similarly, the cost of one hors d'oeuvre for Ann is 2 ounces of vegetables. Bob's costs are the inverse of Ann's—an ounce of vegetables costs him two hors d'oeuvres.

Figure 6.3 Joint PPF

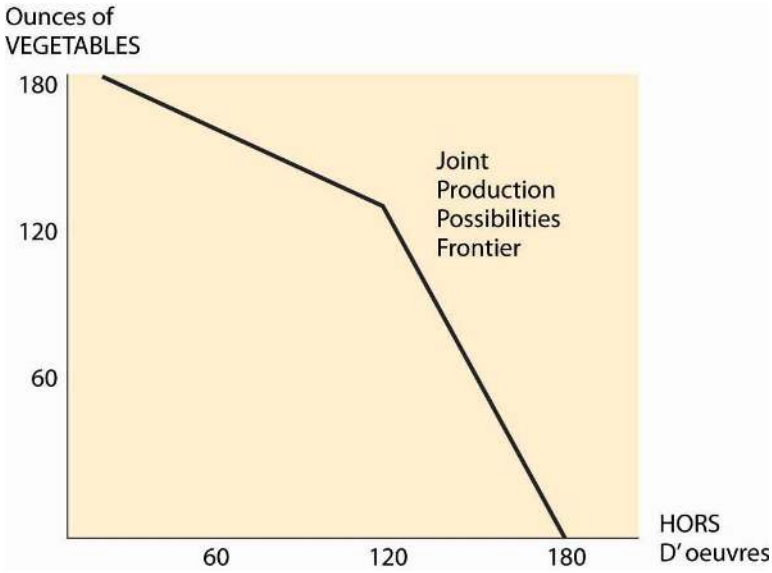


Figure 6.3. Together, they can make 180 of one and none of the other. If Bob makes only hors d'oeuvres, and Ann makes only chopped vegetables, they will have 120 of each. With fewer than 120 ounces of vegetables, the opportunity cost of vegetables is Ann's, and is thus half an hors d'oeuvre; but if more than 120 are needed, then the opportunity cost jumps to two.

Now change the hypothetical slightly. Suppose that Bob and Ann are putting on separate dinner parties, each of which will feature chopped vegetables and hors d'oeuvres in equal portions. By herself, Ann can only produce 40 ounces of vegetables and 40 hors d'oeuvres if she must produce equal portions. She accomplishes this by spending 20 minutes on vegetables and 40 minutes on hors d'oeuvres. Similarly, Bob can produce 40 of each, but by using the reverse allocation of time.

By working together, they can collectively have more of both goods. Ann specializes in producing vegetables, and Bob specializes in producing hors d'oeuvres. This yields 120 units of each, which they can split equally to have 60 of each. By specializing in the activity in which they have lower cost, Bob and Ann can jointly produce more of each good.

Moreover, Bob and Ann can accomplish this by trading. At a "one for one" price, Bob can produce 120 hors d'oeuvres, and trade 60 of them for 60 ounces of vegetables. This is better than producing the vegetables himself, which netted him only 40 of each. Similarly, Ann produces 120 ounces of vegetables and trades 60 of them for 60 hors d'oeuvres. This trading makes them both better off.

The gains from specialization are potentially enormous. The grandfather of economics, Adam Smith, wrote about specialization in the manufacture of pins:

One man draws out the wire; another straights it; a third cuts it; a fourth points it; a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations; to put it on is a peculiar business; to whiten the pins is another; it is even a trade by itself to put them into the paper; and the important business of making a pin is, in this manner, divided into about eighteen distinct operations, which, in some manufactories, are all performed by distinct hands, though in others the same man will sometimes perform two or three of them.Adam Smith, An Inquiry into the Nature and Causes of the Wealth of Nations, originally published in 1776, released by the Gutenberg project, 2002.

Smith goes on to say that skilled individuals could produce at most 20 pins per day acting alone; but that, with specialization, 10 people could produce 48,000 pins per day, 240 times as many pins per capita.

Key Takeaways

- The production possibilities set is the collection of "feasible outputs" of an individual or group.
- The boundary of the production possibilities set is known as the production possibilities frontier.
- The principle of diminishing marginal returns implies that the production possibilities frontier is concave toward the origin, which is equivalent to increasing opportunity cost.
- Efficiencies created by specialization create the potential for gains from trade.

EXERCISES

1. The Manning Company has two factories, one that makes roof trusses and one that makes cabinets. With m workers, the roof factory produces m trusses per day. With n workers, the cabinet plant produces 5 n . The Manning Company has 400 workers to use in the two factories. Graph the production possibilities frontier. (Hint: Let T be the number of trusses produced. How many workers are used to make trusses?)
2. Alarm & Tint, Inc., has 10 workers working a total of 400 hours per week. Tinting takes 2 hours per car. Alarm installation is complicated, however, and performing A alarm installations requires A^2 hours of labor. Graph Alarm & Tint's production possibilities frontier for a week.
3. Consider two consumers and two goods, x and y. Consumer 1 has utility $u_1(x_1, y_1) = x_1 + y_1$ and consumer 2 has utility $u_2(x_2, y_2) = \min\{x_2, y_2\}$. Consumer 1 has an endowment of (1, 1/2) and consumer 2's endowment is (0, 1/2).
    1. Draw the Edgeworth box for this economy.
    2. Find the contract curve and the individually rational part of it. (You should describe these in writing and highlight them in the Edgeworth box.)
    3. Find the prices that support an equilibrium of the system and the final allocation of goods under those prices.

    For Questions 4 to 7, consider an orange juice factory that uses, as inputs, oranges and workers. If the factory uses x pounds of oranges and y workers per hour, it produces gallons of orange juice; $T = 20x^{0.25}y^{0.5}$.

4. Suppose oranges cost $1 and workers cost $10. What relative proportion of oranges and workers should the factory use?
5. Suppose a gallon of orange juice sells for $1. How many units should be sold, and what is the input mix to be used? What is the profit?
6. Generalize the previous exercise for a price of p dollars per gallon of orange juice.
7. What is the supply elasticity?

This page titled 6.1: Production Possibilities Frontier is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

6.2: Comparative and Absolute Advantage 6.2: Comparative and Absolute AdvantageEdit Update titlecancel

Learning Objectives

- Who can produce more?
- How does that relate to cost?
- Can a nation be cheaper on all things?

Ann produces chopped vegetables because her opportunity cost of producing vegetables, at half of one hors d'oeuvre, is lower than Bob's. When one good has a lower opportunity cost over another, it is said to have a **comparative advantage**. That is, Ann gives up less to produce chopped vegetables than Bob, so in comparison to hors d'oeuvres, she has an advantage in the production of vegetables. Since the cost of one good is the amount of another good forgone, a comparative advantage in one good implies a **comparative disadvantage**—a higher opportunity cost—in another. If you are better at producing butter, you are necessarily worse at something else—and, in particular, the thing you give up less of to get more butter.

To illustrate this point, let's consider another party planner. Charlie can produce one hors d'oeuvre or 1 ounce of chopped vegetables per minute. His production is strictly less than Ann's; that is, his production possibilities frontier lies inside of Ann's. However, he has a comparative advantage over Ann in the production of hors d'oeuvres because he gives up only 1 ounce of vegetables to produce an hors d'oeuvre, while Ann must give up 2 ounces of vegetables. Thus, Ann and Charlie can still benefit from trade if Bob isn't around.

When one production possibilities frontier lies outside another, the larger is said to have an **absolute advantage**—it can produce more of all goods than the smaller. In this case, Ann has an absolute advantage over Charlie—she can, by herself, have more—but not over Bob. Bob has an absolute advantage over Charlie, too; but again, not over Ann.

Diminishing marginal returns implies that the more of a good that a person produces, the higher the cost is (in terms of the good given up). That is to say, diminishing marginal returns means that supply curves slope upward; the marginal cost of producing more is increasing in the amount produced.

Trade permits specialization in activities in which one has a comparative advantage. Moreover, whenever opportunity costs differ, potential gains from trade exist. If Person 1 has an opportunity cost of c_1 of producing good x (in terms of y, that is, for each unit of x that Person 1 produces, Person 1 gives up c_1 units of y), and Person 2 has an opportunity cost of c_2, then there are gains from trade whenever c_1 is not equal to c_2 and neither party has specialized.If a party specialized in one product, it is a useful convention to say that the marginal cost of that product is now infinite, since no more can be produced. Suppose $c_1 < c_2$. Then by having Person 1 increase the production of x by Δ, $c_1\Delta$ less of the good y is produced. Let Person 2 reduce the production of x by Δ so that the production of x is the same. Then there is $c_2\Delta$ units of y made available, for a net increase of $(c_2 - c_1)\Delta$. The net changes are summarized in Table 6.1.

Table 6.1 Construction of the gains from trade

	1	2	Net Change
Change in x	$+\Delta$	$-\Delta$	0
Change in y	$-c_1\Delta$	$c_2\Delta$	$(c_2 - c_1)\Delta$

Whenever opportunity costs differ, there are gains from reallocating production from one producer to another, gains which are created by having the low-cost producers produce more, in exchange for greater production of the other good by the other producer, who is the low-cost producer of this other good. An important aspect of this reallocation is that it permits production of more of all goods. This means that there is little ambiguity about whether it is a good thing to reallocate production—it just means that we have more of everything we want.If you are worried that more production means more pollution or other bad things, rest assured. Pollution is bad, so we enter the negative of pollution (or environmental cleanliness) as one of the goods we would like to have on hand. The reallocation dictated by differences in marginal costs produces more of all goods. Now with this said, we have no reason to believe that the reallocation will benefit everyone—there may be winners and losers.

How can we guide the reallocation of production to produce more goods and services? It turns out that, under some circumstances, the price system does a superb job of creating efficient production. The price system posits a price for each good or service, and anyone can sell at the common price. The insight is that such a price induces efficient production. To see this, suppose we have a price *p*, which is the number of units of *y* that one has to give to get a unit of *x*. (Usually prices are in currency, but we can think of them as denominated in goods, too.) If I have a cost *c* of producing *x*, which is the number of units of *y* that I lose to obtain a unit of *x*, I will find it worthwhile to sell *x* if $p > c$, because the sale of a unit of *x* nets me $p - c$ units of *y*, which I can either consume or resell for something else I want. Similarly, if $c > p$, I would rather buy *x* (producing *y* to pay for it). Either way, only producers with costs less than *p* will produce *x*, and those with costs greater than *p* will purchase *x*, paying for it with *y*, which they can produce more cheaply than its price. (The price of *y* is $1/p$—that is, the amount of *x* one must give to get a unit of *y*.)

Thus, a price system, with appropriate prices, will guide the allocation of production to ensure the low-cost producers are the ones who produce, in the sense that there is no way of reallocating production to obtain more goods and services.

Key Takeaways

- A lower opportunity cost creates a comparative advantage in production.
- A comparative advantage in one good implies a comparative disadvantage in another.
- It is not possible to have a comparative disadvantage in all goods.
- An absolute advantage means the ability to produce more of all goods.
- Diminishing marginal returns implies that the more of a good that a person produces, the higher is the cost (in terms of the good given up). That is to say, diminishing marginal returns means that supply curves slope upward; the marginal cost of producing more is increasing in the amount produced.
- Trade permits specialization in activities in which one has a comparative advantage.
- Whenever opportunity costs differ, potential gains from trade exist.
- Trade permits production of more of all goods.
- A price system, with appropriate prices, will guide the allocation of production to ensure the low-cost producers are the ones who produce, in the sense that there is no way of reallocating production to obtain more goods and services.

EXERCISES

1. Graph the joint production possibilities frontier for Ann and Charlie, and show that collectively they can produce 80 of each if they need the same number of each product. (Hint: First show that Ann will produce some of both goods by showing that, if Ann specializes, there are too many ounces of vegetables. Then show that, if Ann devotes *x* minutes to hors d'oeuvres, $60 + x = 2(60 - x)$.)
2. Using Manning's production possibilities frontier in Exercise 1 in Section 6.1, compute the marginal cost of trusses in terms of cabinets.
3. Using Alarm & Tint's production possibilities frontier in Exercise 2 in Section 6.1, compute the marginal cost of alarms in terms of window tints.

This page titled 6.2: Comparative and Absolute Advantage is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

6.3: Factors of Production 6.3: Factors of ProductionEdit Update titlecancel

Learning Objectives

- How does the abundance or rarity of inputs to production affect the advantage of nations?

Production possibilities frontiers provide the basis for a rudimentary theory of international trade. To understand the theory, it is first necessary to consider that there are fixed and mobile factors. **Factors of production** is jargon for inputs to the production process. Labor is generally considered a fixed factor because most countries don't have borders that are wide open to immigration, although of course some labor moves across international borders. Temperature, weather, and land are also fixed—Canada is a high-cost citrus grower because of its weather. There are other endowments that could be exported but are expensive to export because of transportation costs, including water and coal. Hydropower—electricity generated from the movement of water—is cheap and abundant in the Pacific Northwest; and, as a result, a lot of aluminum is smelted there because aluminum smelting requires lots of electricity. Electricity can be transported, but only with losses (higher costs), which gives other regions a disadvantage in the smelting of aluminum. Capital is generally considered a mobile factor because plants can be built anywhere, although investment is easier in some environments than in others. For example, reliable electricity and other inputs are necessary for most factories. Moreover, comparative advantage may arise from the presence of a functioning legal system, the enforcement of contracts, and the absence of bribery. This is because enforcement of contracts increase the return on investment by increasing the probability that the economic return to investment isn't taken by others.

Fixed factors of production are factors that are not readily moved and thus give particular regions a comparative advantage in the production of some kinds of goods and not in others. Europe, the United States, and Japan have a relative abundance of highly skilled labor and have a comparative advantage in goods requiring high skills like computers, automobiles, and electronics. Taiwan, South Korea, Singapore, and Hong Kong have increased the available labor skills and now manufacture more complicated goods like DVDs, computer parts, and the like. Mexico has a relative abundance of middle-level skills, and a large number of assembly plants operate there, as well as clothing and shoe manufacturers. Lower-skilled Chinese workers manufacture the majority of the world's toys. The skill levels of China are rising rapidly.

The basic model of international trade, called **Ricardian theory**, was first described by David Ricardo (1772–1823). It suggests that nations, responding to price incentives, will specialize in the production of goods in which they have a comparative advantage, and purchase the goods in which they have a comparative disadvantage. In Ricardo's description, England has a comparative advantage of manufacturing cloth and Portugal similarly in producing wine, leading to gains from trade from specialization.

The Ricardian theory suggests that the United States, Canada, Australia, and Argentina should export agricultural goods, especially grains that require a large land area for the value generated (they do). It suggests that complex technical goods should be produced in developed nations (they are) and that simpler products and natural resources should be exported by the lesser-developed nations (they are). It also suggests that there should be more trade between developed and underdeveloped nations than between developed and other developed nations. The theory falters on this prediction—the vast majority of trade is between developed nations. There is no consensus for the reasons for this, and politics plays a role—the North American Free Trade Act (NAFTA) vastly increased the volume of trade between the United States and Mexico, for example, suggesting that trade barriers may account for some of the lack of trade between the developed and the underdeveloped world. Trade barriers don't account for the volume of trade between similar nations, which the theory suggests should be unnecessary. Developed nations sell each other such products as mustard, tires, and cell phones, exchanging distinct varieties of goods they all produce.

Key Takeaways

- The term "factors of production" is jargon for inputs to the production process.
- Labor is generally considered a fixed or immobile factor because most countries don't have borders that are wide open to immigration. Temperature, weather, and land are also fixed factors.
- Fixed factors of production give particular regions a comparative advantage in the production of some kinds of goods, and not in others.
- The basic model of international trade, known as the Ricardian theory, suggests that nations, responding to price incentives, will specialize in the production of goods in which they have a comparative advantage, and purchase the goods in which they have a comparative disadvantage.

This page titled 6.3: Factors of Production is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

6.4: International Trade 6.4: International TradeEdit Update titlecancel

Learning Objectives

- How does trade affect domestic prices for inputs and goods and services?

The Ricardian theory emphasizes that the relative abundance of particular factors of production determines comparative advantage in output, but there is more to the theory. When the United States exports a computer to Mexico, American labor, in the form of a physical product, has been sold abroad. When the United States exports soybeans to Japan, American land (or at least the use of American land for a time) has been exported to Japan. Similarly, when the United States buys car parts from Mexico, Mexican labor has been sold to the United States; and when Americans buy Japanese televisions, Japanese labor has been purchased. The goods that are traded internationally embody the factors of production of the producing nations, and it is useful to think of international trade as directly trading the inputs through the incorporation of inputs into products.

If the set of traded goods is broad enough, **factor price equalization** predicts that the value of factors of production should be equalized through trade. The United States has a lot of land, relative to Japan; but by selling agricultural goods to Japan, it is as if Japan has more land, by way of access to U.S. land. Similarly, by buying automobiles from Japan, it is as if a portion of the Japanese factories were present in the United States. With inexpensive transportation, the trade equalizes the values of factories in the United States and Japan, and also equalizes the value of agricultural land. One can reasonably think that soybeans are soybeans, wherever they are produced, and that trade in soybeans at a common price forces the costs of the factors involved in producing soybeans to be equalized across the producing nations. The purchase of soybeans by the Japanese drives up the value of American land and drives down the value of Japanese land by giving an alternative to its output, leading toward equalization of the value of the land across the nations.

Factor price equalization was first developed by Paul Samuelson (1915–) and generalized by Eli Heckscher (1879–1952) and Bertil Ohlin (1899–1979). It has powerful predictions, including the equalization of wages of equally skilled people after free trade between the United States and Mexico. Thus, free trade in physical goods should equalize the price of such items as haircuts, land, and economic consulting in Mexico City and New York City. Equalization of wages is a direct consequence of factor price equalization because labor is a factor of production. If economic consulting is cheap in Mexico, trade in goods embodying economic consulting—boring reports, perhaps—will bid up the wages in the low-wage area and reduce the quantity in the high-wage area.

An even stronger prediction of the theory is that the price of water in New Mexico should be the same as in Minnesota. If water is cheaper in Minnesota, trade in goods that heavily use water—for example, paper—will tend to bid up the value of Minnesota water while reducing the premium on scarce New Mexico water.

It is fair to say that if factor price equalization works fully in practice, it works very, very slowly. Differences in taxes, tariffs, and other distortions make it a challenge to test the theory across nations. On the other hand, within the United States, where we have full factor mobility and product mobility, we still have different factor prices—electricity is cheaper in the Pacific Northwest. Nevertheless, nations with a relative abundance of capital and skilled labor export goods that use these intensively; nations with a relative abundance of land export land-intensive goods like food; nations with a relative abundance of natural resources export these resources; and nations with an abundance of low-skilled labor export goods that make intensive use of this labor. The reduction of trade barriers between such nations works like Ann and Bob's joint production of party platters. By specializing in the goods in which they have a comparative advantage, there is more for all.

Key Takeaways

- Goods that are traded internationally embody the factors of production of the producing nations.
- It is useful to think of international trade as directly trading the inputs through the incorporation of inputs into products.
- If the set of traded goods is broad enough, the value of factors of production should be equalized through trade. This prediction is known as factor price equalization.

This page titled 6.4: International Trade is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

7: Externalities

7.1: External Effects
7.2: Pigouvian Taxes
7.3: Quotas
7.4: Tradable Permits and Auctions
7.5: Coasian Bargaining
7.6: Fishing and Extinction

This page titled 7: Externalities is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

7.1: External Effects 7.1: External EffectsEdit Update titlecancel

When the person sitting next to you lights up a cigarette, he gets nicotine and the cigarette company gets some of his money. You just suffer, with no compensation. If your neighbor's house catches fire because he fell asleep with that cigarette burning in his hand, your house may burn to the ground. The neighbor on the other side who plays very loud music late into the night—before your big economics test—enjoys the music, and the music distributors gets his money. You flunk out of college and wind up borrowing $300,000 to buy a taxi medallion. Drunk drivers, cell phones ringing in movie theaters, loud automobiles, polluted air, and rivers polluted to the point that they catch fire, like Cleveland's Cuyahoga did, are all examples where a transaction between two parties harmed other people. These are "external effects."

But external effects are not necessarily negative. The neighbor who plants beautiful flowers in her yard brightens your day. Another person's purchase of an electric car reduces the smog you breathe. Your neighbor's investment in making his home safe from fire conveys a safety advantage to you. Indeed, even your neighbor's investment in her own education may provide an advantage to you—you may learn useful things from your neighbor. Inventions and creations, whether products or poetry, produce value for others. The creator of a poem or a mathematical theorem provides a benefit to others.

Learning Objectives

- How can society stop people from doing annoying things and encourage them to do pleasing things?

These effects are called external effects, or externalities. An **externality** is any effect on people not involved in a particular transaction. Pollution is the classic example. When another person buys and smokes cigarettes, there is a transaction between the cigarette company and the smoker. But if you are sitting near the smoker, you are an affected party who is not directly compensated from the transaction, at least before taxes were imposed on cigarettes. Similarly, you pay nothing for the benefits you get from viewing your neighbor's flowers, nor is there a direct mechanism to reward your neighbor for her efforts.

Externalities will generally cause competitive markets to behave inefficiently from a social perspective, absent a mechanism to involve all the affected parties. Without such a mechanism, the flower planter will plant too few beautiful flowers, for she has no reason to take account of your preferences in her choices. The odious smoker will smoke too much and too close to others, and the loud neighbor will play music much too late into the night. Externalities create a **market failure**—that is, a situation where a competitive market does not yield the socially efficient outcome.

Education is viewed as creating an important positive externality. Education generates many externalities, including more—and better—employment, less crime, and fewer negative externalities of other kinds. It is widely believed that educated voters elect better politicians.This is a logical proposition, but there is scant evidence in favor of it. There is evidence that educated voters are more likely to vote, but little evidence that they will vote for better candidates. Educated individuals tend to make a society wealthy, an advantage to all of society's members. As a consequence, most societies subsidize education in order to promote it.

A major source of externalities arises in communicable diseases. Your vaccination not only reduces the likelihood that you will contract a disease but also makes it less likely that you will infect others with the disease.

Let's consider pollution as a typical example. A paper mill produces paper, and a bad smell is an unfortunate by-product of the process. Each ton of paper produced increases the amount of bad smells produced. The paper mill incurs a marginal cost, associated with inputs like wood and chemicals and water. For the purposes of studying externalities, we will refer to the paper mill's costs as a **private cost**, the cost borne by the supplier (in this case, the paper mill itself). In addition, there are **external costs**, which are the costs borne by third parties, that arise in this case from the smell. Adding the private costs and the external costs yield the total costs for all parties, or the **social costs**. These costs, in their marginal form, are illustrated in Figure 7.1.

Figure 7.1 A negative externality

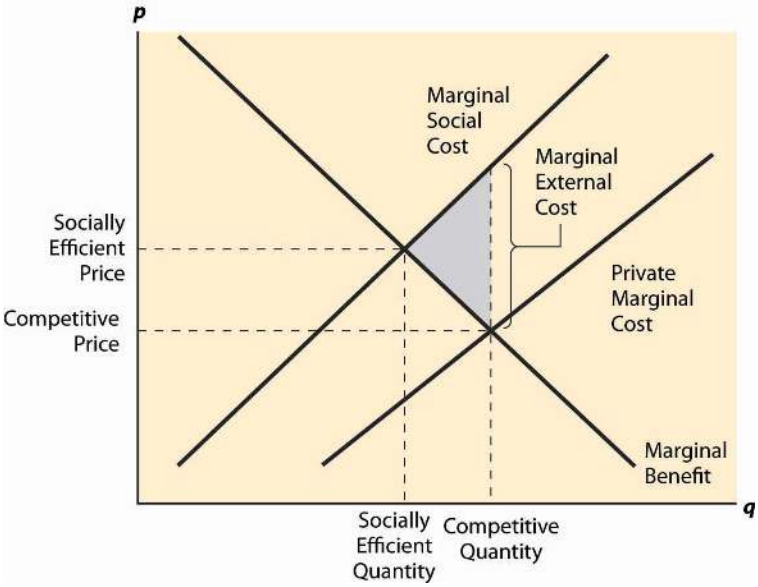


Figure 7.1, the demand has been labeled "marginal benefit," for reasons that will become apparent; but it is at this point just the standard demand, the marginal value of the product. The paper mill's costs have been labeled marginal private cost to reflect the fact that these costs are only the mill's costs and don't include the cost of the bad smell imposed on others. The marginal social cost is obtained by adding the marginal external cost to the marginal private cost. The marginal external cost isn't graphed in the figure; but the size of it is illustrated at one quantity, and it is generally the difference between marginal social cost and marginal private cost.

Left to its own devices, the paper market would equate the marginal private cost and the marginal benefit to produce the competitive quantity sold at the competitive price. Some of these units—all of those beyond the quantity labeled "Socially Efficient Quantity"—are bad from a social perspective: They cost more to society than they provide in benefits. This is because the social cost of these units includes pollution, but paper buyers have no reason to worry about pollution or even to know that it is being created in the process of manufacturing paper.

The deadweight loss of these units is shown as a shaded triangle in the figure. The loss arises because the marginal social cost of the units exceeds the benefit, and the difference between the social cost and the benefits yields the loss to society. This is a case where too much is produced because the market has no reason to account for all the costs; some of the costs are borne by others.

Figure 7.2 External costs and benefits

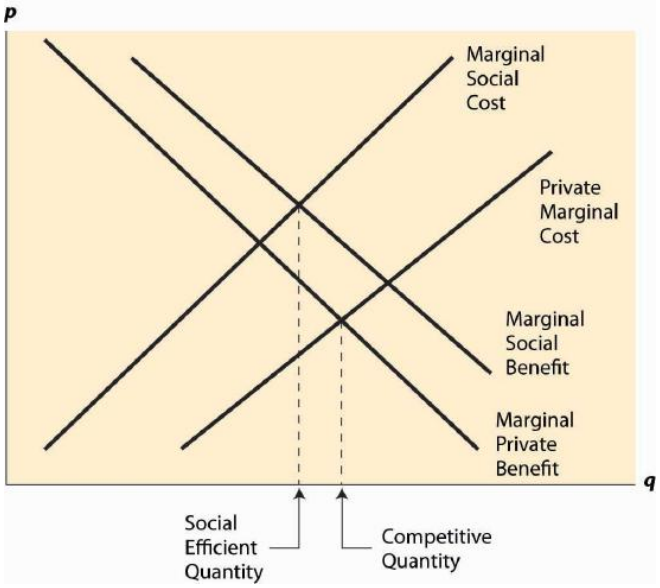


Figure 7.2. Streetlights are an example of a product that produces both externalities: Most of us like lighted streets, but they are terrible for astronomers. Similarly, large highways produce benefits for commuters and yet harm nearby residents.

The marginal **private benefit**—the benefit obtained by the buyer—and the marginal private cost give the demand and supply of a competitive market, and hence the competitive quantity results from the intersection of these two. The marginal social benefit and the marginal social cost give the value and cost from a social perspective; equating these two generates the socially efficient outcome. This can be either greater or less than the competitive outcome depending on which externality is larger.

Consider a town on a scenic bay that is filled with lobsters. The town members collect and eat the lobsters, and over time the size of the lobsters collected falls, until they are hardly worth searching for. This situation persists indefinitely. Few large lobsters are caught, and it is barely worth one's time attempting to catch them. This sort of overuse of a resource due to lack of ownership is known as the **tragedy of the commons**.

The tragedy of the commons is a problem with a **common resource** shared by many people—in this case, the lobster bay. Catching lobsters creates an externality by lowering the productivity of other lobster catchers. The externality leads to overfishing, since individuals don't take into account the negative effect they have on each other, ultimately leading to a nearly useless resource and potentially driving the lobsters to extinction. As a consequence, the lobster catching is usually regulated.

Key Takeaways

- An externality is any effect on people not involved in a particular transaction.
- Pollution is the classic negative externality.
- Externalities will generally cause competitive markets to behave inefficiently from a social perspective. Externalities create a market failure—that is, a competitive market does not yield the socially efficient outcome.
- Education is viewed as creating an important positive externality.
- A major source of externalities arises in communicable diseases. Your vaccination not only reduces the likelihood that you will contract a disease but also makes it less likely that you will infect others with the disease.
- Private costs are those borne by the parties to a transaction; external costs are costs borne by others not party to the transaction; and social costs represent the sum of private and external costs.
- Private benefits are those enjoyed by the parties to a transaction; external benefits are those enjoyed by others not party to the transaction; and social benefits represent the sum of private and external benefits.
- Demand is marginal private benefit; supply is marginal private cost.
- The social optimum arises at the quantity where marginal social benefit equals marginal social cost. A different quantity than the social optimum creates a deadweight loss equal to the difference of marginal social benefit and cost.
- The tragedy of the commons is the overuse of a resource arising because users ignore the harmful effects of their use on other users.

EXERCISES

1. A child who is vaccinated against polio is more likely to contract polio (from the vaccine) than an unvaccinated child. Does this fact imply that programs forcing vaccination of schoolchildren are ill-advised? Include with your answer a diagram illustrating the negative marginal benefit of vaccination, and use the horizontal axis to represent the proportion of the population vaccinated.
2. The total production from an oil field generally depends on the rate at which the oil is pumped, with faster rates leading to lower total production but earlier production. Suppose two different producers can pump from the field. Illustrate—using an externality diagram where the horizontal axis is the rate of production for one of the producers—the difference between the socially efficient outcome and the equilibrium outcome. Like many other states, Texas's law requires that, when multiple people own land over a single oil field, the output is shared among the owners, with each owner obtaining a share equal to a proportion of the field under their land. This process is called unitization. Does it solve the problem of externalities in pumping and yield an efficient outcome? Why or why not?
3. Imagine that many students are bothered by loud music playing at 7 a.m. near their dorm. Using economic analysis, how would you improve the situation?
4. A local community uses revenue from its property taxes to build an expressway. Explain how this could give rise to free riders and how to solve the free-rider problem.

This page titled 7.1: External Effects is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

7.2: Pigouvian Taxes 7.2: Pigouvian TaxesEdit Update titlecancel

Learning Objectives

- Can society regulate annoying behavior with taxes?

Arthur Cecil Pigou (1877–1959) proposed a solution to the problem of externalities that has become a standard approach. This simple idea is to impose a per-unit tax on a good, thereby generating negative externalities equal to the marginal externality at the socially efficient quantity. This is known as a **Pigouvian tax**. Thus, if at the socially efficient quantity, the marginal external cost is $1, then a $1 per-unit tax would lead to the right outcome. This is illustrated in Figure 7.3.

The tax that is added is the difference, at the socially efficient quantity, between the marginal social cost and the marginal private cost, which equals the marginal external cost. The tax level need not equal the marginal external cost at other quantities, and the figure reflects a marginal external cost that is growing as the quantity grows. Nevertheless, the new supply curve created by the addition of the tax intersects demand (the marginal benefit) at the socially efficient quantity. As a result, the new competitive equilibrium, taking account of the tax, is efficient.

The case of a positive externality is similar. Here, a subsidy is needed to induce the efficient quantity. It is left as an exercise.

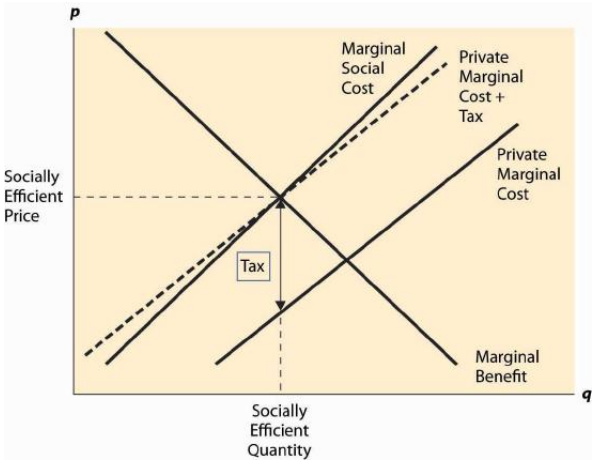


Figure 7.3 The Pigouvian tax

Taxes and subsidies are fairly common instruments to control externalities. We subsidize higher education with state universities, and the federal government provides funds for research and limited funds for the arts. Taxes on cigarettes and alcoholic beverages are used to discourage these activities, perhaps because smoking and drinking alcoholic beverages create negative externalities. (Cigarettes and alcohol also have inelastic demands, which make them good candidates for taxation since there is only a small distortion of the quantity.) However, while important in some arenas, taxes and subsidies are not the most common approach to regulation of externalities.

Key Takeaways

- A Pigouvian tax is a per-unit tax on a good, thereby generating negative externalities equal to the marginal externality at the socially efficient quantity.
- Imposition of a Pigouvian tax leads to a competitive equilibrium, taking account of the tax, which is efficient.
- In the case of a positive externality, a subsidy can be used to obtain efficiency.
- Taxes and subsidies are fairly common instruments to control externalities.

EXERCISES

1. Identify the tax revenue produced by a Pigouvian tax in Figure 7.3. What is the relationship between the tax revenue and the damage produced by the negative externality? Is the tax revenue sufficient to pay those damaged by the external effect an amount equal to their damage? (Hint: Is the marginal external effect increasing or decreasing?)
2. Identify, by using a diagram, the Pigouvian subsidy needed to induce the efficient quantity in the case of a positive externality. When is the subsidy expended smaller than the total external benefit?

3. Use the formulae for estimating the effect of a tax on quantity to deduce the size of the tax needed to adjust for an externality when the marginal social cost is twice the marginal private cost.

This page titled 7.2: Pigouvian Taxes is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

7.3: Quotas 7.3: QuotasEdit Update titlecancel

Learning Objectives

- Can society regulate annoying behavior by just telling people what to do?

The Pigouvian tax and subsidy approach to dealing with externalities has several problems. First, it requires knowing the marginal value or cost of the external effect, and this may be a challenge to estimate. Second, it requires the imposition of taxes and permits the payment of subsidies, which encourages what might be politely termed as "misappropriation of funds." That is, once a government agency is permitted to tax some activities and subsidize others, there will be a tendency to tax things people in the agency don't like and subsidize "pet" projects, using the potential for externalities as an excuse rather than a real reason. U.S. politicians have been especially quick to see positive externalities in oil, cattle, and the family farm—externalities that haven't been successfully articulated. (The Canadian government, in contrast, sees externalities in filmmaking and railroads.)

An alternative to the Pigouvian tax or subsidy solution is to set a quota, which is a limit on the activity. Quotas can be maxima or minima, depending on whether the activity generates negative or positive externalities. We set maximum levels on many pollutants rather than tax them, and ban some activities, like lead in gasoline or paint, or chlorofluorocarbons outright (a quota equal to zero). We set maximum amounts on impurities, like rat feces, in foodstuffs. We impose minimum educational attainment (eighth grade or age 16, whichever comes first), minimum age to drive, and minimum amount of rest time for truck drivers and airline pilots. A large set of regulations governs electricity and plumbing, designed to promote safety, and these tend to be "minimum standards." Quotas are a much more common regulatory strategy for dealing with externalities than taxes and subsidies.

The idea behind a quota is to limit the quantity to the efficient level. If a negative externality in pollution means our society pollutes too much, then impose a limit or quantity restriction on pollution. If the positive externality of education means individuals in our society receive too little education from the social perspective, force them to go to school.

As noted, quotas have the advantage of addressing the problem without letting the government spend more money, limiting the government's ability to misuse funds. On the other hand, quotas have the problem of identifying who should get the quota; quotas will often misallocate the resource. Indeed, a small number of power plants account for almost half of the man-made sulfur dioxide (SO_2) pollution emitted into the atmosphere, primarily because these plants historically emitted a lot of pollution and their pollution level was set by their historical levels. Quotas tend to harm new entrants compared to existing firms and discourage the adoption of new technology. Indeed, the biggest polluters must stay with old technology in order to maintain their right to pollute.

Key Takeaways

- An alternative to the Pigouvian tax or subsidy solution is to set a quota. Quotas can be maxima or minima, depending on whether the activity generates negative or positive externalities.
- Quotas are a much more common regulatory strategy for dealing with externalities than taxes and subsidies.
- The goal of a quota is to limit the quantity to the efficient level.
- Quotas tend to harm new entrants compared to existing firms and discourage the adoption of new technology.

EXERCISES

1. If a quota is set to the socially efficient level, how does the value of a quota right compare to the Pigouvian tax?
2. Speeding (driving fast) creates externalities by increasing the likelihood and severity of automobile accidents, and most countries put a limit on speed; but one could instead require fast drivers to buy a permit to speed. Discuss the advantages and disadvantages of "speeding permits."

This page titled 7.3: Quotas is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

7.4: Tradable Permits and Auctions 7.4: Tradable Permits and AuctionsEdit Update titlecancel

Learning Objectives

- Is there a better way to regulate annoying behavior than either taxes or quotas?

A solution to inefficiencies in the allocation of quota rights is to permit trading them. **Tradable permits** are quotas for pollution that can be exchanged to create a market in the right to pollute, and thereby create a tax on polluting. The emission of pollution requires the purchase of permits to pollute, and the price of these permits represents a tax on pollution. Thus, tradable permits represent a hybrid of a quota system and a Pigouvian taxation system—a quota determines the overall quantity of pollution as in a quota system, determining the supply of pollution rights, but the purchase of pollution rights acts like a tax on pollution, a tax whose level is determined by the quota supply and demand.

Figure 7.4 SO_2 permit prices



Figure 7.4 shows the price of sulfur dioxide permits over the past decade.Source: Environmental Protection Agency, July 22, 2004, www.epa.gov/airmarkets/trading/so2market/alprices.html Each permit conveys the right to emit one ton of sulfur dioxide into the air. The overall pollution level is being reduced over time, which accounts for some of the increase in prices. These prices represent significant taxes on large polluters, as a coal-fired power plant using coal with high sulfur content can annually produce as much as 200,000 tons of sulfur dioxide.

The major advantage of a tradable permits system is that it creates the opportunity for efficient exchange—one potential polluter can buy permits from another, leaving the total amount of pollution constant. Such exchange is efficient because it uses the pollution in a manner creating the highest value, eliminating a bias toward “old” sources. Indeed, a low-value polluter might sell its permits and just shut down if the price of pollution was high enough.

A somewhat unexpected advantage of tradable permits has been the purchase of permits by environmental groups like the Sierra Club. Environmental groups can buy permits and then not exercise them, as a way of cleaning the air. In this case, the purchase of the permits creates a major positive externality on the rest of society, since the environmental group expends its own resources to reduce pollution of others.

Tradable permits offer the advantages of a taxation scheme—efficient use of pollution—without needing to estimate the social cost of pollution directly. This is especially valuable when the strategy is to set a quantity equal to the current quantity, and then gradually reduce the quantity in order to reduce the effects of the pollution. The price of permits can be a very useful instrument in assessing the appropriate time to reduce the quantity, since high permit prices, relative to likely marginal external costs, suggest that the quantity of the quota is too low, while low prices suggest that the quantity is too large and should be reduced.

Key Takeaways

- A solution to inefficiencies in the allocation of quota rights is to permit trading them.
- Tradable permits represent a hybrid of a quota system and a Pigouvian taxation system. The quota determines the overall quantity of pollution, while the purchase of pollution rights acts like a tax on pollution.
- The United States has permitted the trading of permits for some pollutants, like sulfur dioxide.

- The major advantage of a tradable permits system is that it creates the opportunity for efficient exchange.
- A somewhat unexpected advantage of tradable permits has been the purchase of permits by environmental groups, as a way of buying cleaner air.
- Tradable permits offer the advantages of a taxation scheme—efficient use of pollution—without needing to estimate the social cost of pollution directly.
- The price of permits can be a very useful instrument in assessing the appropriate time to reduce the quantity, since high permit prices, relative to likely marginal external costs, suggest that the quantity of the quota is too low, while low prices suggest that the quantity is too large and should be reduced.

This page titled 7.4: Tradable Permits and Auctions is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

7.5: Coasian Bargaining 7.5: Coasian BargainingEdit Update titlecancel

Learning Objectives

- Can I just bribe my neighbor to stop being annoying?

The negative externality of a neighbor playing loud music late at night is not ordinarily solved with a tax or with a quota but instead through an agreement. When there aren't many individuals involved, the individuals may be able to solve the problem of externalities without involving a government but through negotiation. This insight was developed by Nobel laureate Ronald Coase (1910–), and is sometimes known as **Coasian bargaining**.

Coase offered the example of a cattle ranch next to a farm. There is a negative externality in that the cattle tend to wander over to the farm and eat the crops, rather than staying on the ranch. What happens next depends on **property rights**, which are the rights that come with ownership.

One of three things might be efficient from a social perspective. It might be efficient to erect a fence to keep the cows away from the crops. It might be efficient to close down the farm. Finally, it might be efficient to close down the ranch, if the farm is valuable enough and if the fence costs more than the value of the ranch.

If the farmer has a right not to have his crops eaten and can confiscate the cows if they wander onto the farm, then the rancher will have an incentive to erect a fence to keep the cows away, if that is the efficient solution. If the efficient solution is to close down the ranch, then the rancher will do that, since the farmer can confiscate the cows if they go over to the farm and it isn't worth building the fence by hypothesis. Finally, if the efficient solution to the externality is to close down the farm, then the rancher will have an incentive to buy the farm in order to purchase the farm's rights so that he can keep the ranch in operation. Since it is efficient to close down the farm only if the farm is worth less than the ranch, there is enough value in operating the ranch to purchase the farm at its value and still have money left over; that is, there are gains from trade from selling the farm to the rancher. In all three cases, if the farmer has the property rights, then the efficient outcome is reached.

Now suppose instead that the rancher has the rights and that the farmer has no recourse if the cows eat his crops. If shutting down the farm is efficient, then the farmer has no recourse but to shut it down. Similarly, if building the fence is efficient, then the farmer will build the fence to protect his crops. Finally, if shutting down the ranch is efficient, the farmer will buy the ranch from the rancher in order to be able to continue to operate the more valuable farm. In all cases, the efficient solution is reached through negotiation.

Coase argued that bargaining can generally solve problems of externalities and that the real problem is ill-defined property rights. If the rancher and the farmer can't transfer their property rights, then the efficient outcome may not arise. In the Coasian view of externalities, if an individual owned the air, air pollution would not be a problem because the owner would charge for the use and wouldn't permit an inefficient level of pollution. The case of air pollution demonstrates some of the limitations of the Coasian approach because ownership of the air, or even the more limited right to pollute into the air, would create an additional set of problems; a case where the cure is likely to be worse than the disease.

Bargaining to solve the problem of externalities is often feasible when a small number of people are involved. When a large number of people are potentially involved, as with air pollution, bargaining is unlikely to be successful in addressing the problem of externalities, and a different approach is required.

Key Takeaways

- When there aren't many individuals involved, the individuals may be able to solve the problem of externalities without involving a government, but through negotiation.
- Nobel laureate Ronald Coase argued that bargaining can generally solve problems of externalities and that the real problem is ill-defined property rights.
- Bargaining to solve the problem of externalities is often feasible when a small number of people are involved. When a large number of people are potentially involved, as with air pollution, bargaining is unlikely to be successful in addressing the problem of externalities.

This page titled 7.5: Coasian Bargaining is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

7.6: Fishing and Extinction 7.6: Fishing and ExtinctionEdit Update titlecancel

Learning Objectives

- Is extinction really an economic phenomenon?
- Why do we overfish?

Consider an unregulated fishing market like the lobster market considered previously, and let S be the stock of fish. The purpose of this example is illustrative of the logic, rather than an exact accounting of the biology of fish populations, but is not unreasonable. Let S be the stock of a particular species of fish. Our starting point is an environment without fishing: How does the fish population change over time? Denote the change over time in the fish population by $S^{\cdot}$ ($S^{\cdot}$ is notation for the derivative with respect to time, notation that dates back to Sir Isaac Newton). We assume that population growth follows the logistic equation $S^{\cdot} = rS(1 - S)$. This equation reflects two underlying assumptions. First, mating and reproduction are proportional to the stock of fish S. Second, survival is proportional to the amount of available resources 1 – S, where 1 is set to be the maximum sustainable population. (Set the units of the number of fish so that 1 is the full population.)

The dynamics of the number of fish are illustrated in Figure 7.5. On the horizontal axis is the number of fish, and on the vertical axis is the change in S. When $S^{\cdot} > 0$, S is increasing over time, and the arrows on the horizontal axis reflect this. Similarly, if $S^{\cdot} < 0$, S is decreasing.

Absent fishing, the value 1 is a stable **steady state** of the fish population, in which the variables stay constant and forces are balanced. It is a steady state because, if $S = 1, S^{\cdot} = 0$; that is, there is no change in the fish population. It is stable because the effect of a small perturbation—S near but not exactly equal to 1—is to return to 1. (In fact, the fish population is very nearly globally stable. Start with any population other than zero and the population returns to 1.)It turns out that there is a closed form solution for the fish population: $S(t) = S(0)S(0) + (1 - S(0))e - rt$.

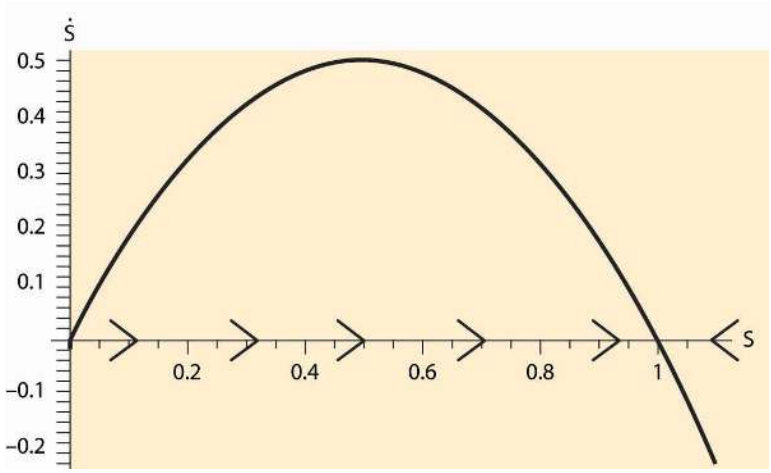


Figure 7.5 Fish population dynamics

Now we introduce a human population and turn to the economics of fishing. Suppose that a boat costs b to launch and operate and that it captures a fixed fraction a of the total stock of fish S; that is, each boat catches aS. Fish sell for a price $p = Q - 1\varepsilon$, where the price arises from the demand curve, which in this case has constant elasticity ε, and Q is the quantity of fish offered for sale. Suppose there are n boats launched; then the quantity of fish caught is Q = naS. Fishers enter the market as long as profits are positive, which leads to zero profits for fishers; that is, $b = (Qn)p(Q)$. This equation makes a company just indifferent to launching an additional boat because the costs and revenues are balanced. These two equations yield two equations in the two unknowns n and $Q : n = Qp(Q)b = 1$ b $Q\varepsilon - 1\varepsilon$, and $Q = naS$. These two equations solve for the number of fish caught, , $Q = ($ as b $)\varepsilon$ and the number of boats, $n = a\varepsilon - 1$ b ε S $\varepsilon - 1$.

Subtracting the capture by humans from the growth in the fish population yields

$$S^{\cdot} = rS(1 - S) - (aSb)\varepsilon$$

Thus, a steady state satisfies $0 = S^{\cdot} = rS(1 - S) - (aSb)\varepsilon$.

Figure 7.6 Fish population dynamics with fishing

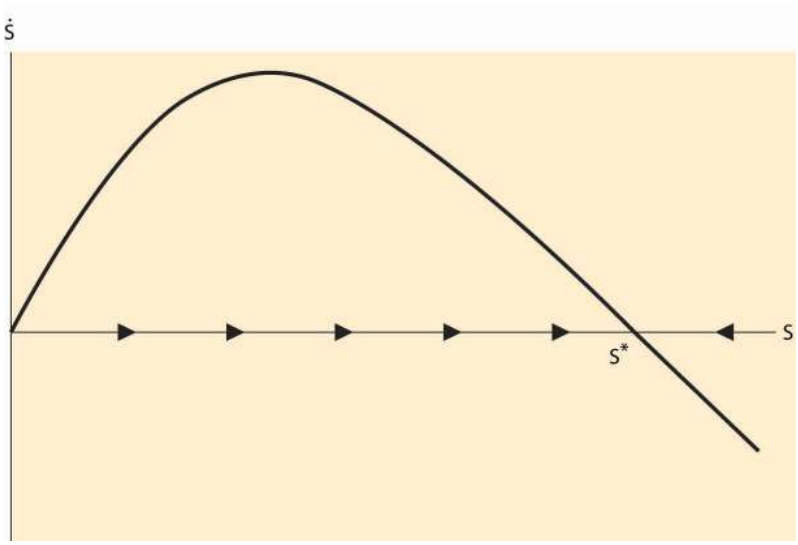


Figure 7.6.

The dark curve represents $\dot{S}$, and thus for S between 0 and the point labeled S^*, $\dot{S}$ is positive and so S is increasing over time. Similarly, to the right of S^*, S is decreasing. Thus, S^* is stable under small perturbations in the stock of fish and is an equilibrium.

We see that if demand for fish is elastic, fishing will not drive the fish to extinction. Even so, fishing will reduce the stock of fish below the efficient level because individual fishers don't take account of the externality they impose—their fishing reduces the stock for future generations. The level of fish in the sea converges to S* satisfying

$$0 = rS^* (1 - S^*) - (aS^* b) \varepsilon$$

In contrast, if demand is inelastic, fishing may drive the fish to extinction. For example, if r = 2 and a = b = 1, and ε = 0.7, extinction is necessary, as is illustrated in Figure 7.7.

Figure 7.7 Fish population dynamics: extinction

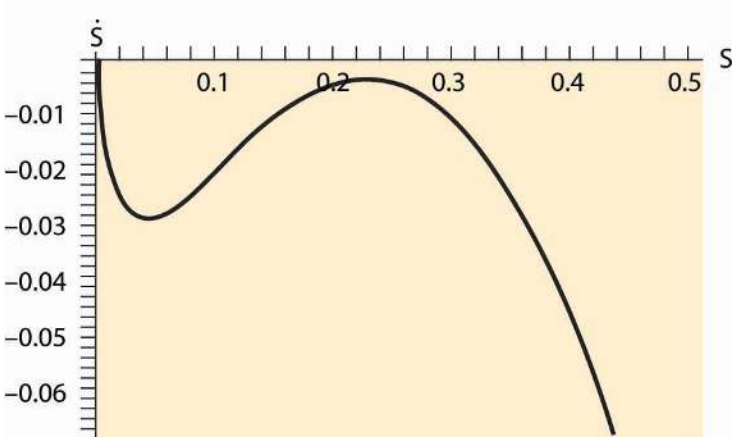


Figure 7.7 shows that, for the given parameters, the net growth of the fish population is negative for every value of the stock S. Thus, the population of fish consistently dwindles. This is a case when the fishing externality (overfishing today reduces the stock of fish tomorrow) has particularly dire consequences. The reason why the elasticity of demand matters is that, with inelastic demand, the fall in the stock of fish increases the price by a large amount (enough so that total revenue rises). This, in turn, increases the number of fishing boats, in spite of the fall in the catch. In contrast, with elastic demand, the number of fishing boats falls as the stock falls, reducing the proportion of fish caught, and thus preventing extinction. We see this for the equation for the number of fishing boats, n= a ε−1 b ε S ε−1 , which reflects the fact that fishing effort rises as the stock falls if and only if demand is inelastic.

It is possible, even with inelastic demand, for there to be a stable fish population. Not all parameter values lead to extinction. Using the same parameters as before, but with ε = 0.9, we obtain a stable outcome as illustrated in Figure 7.8.

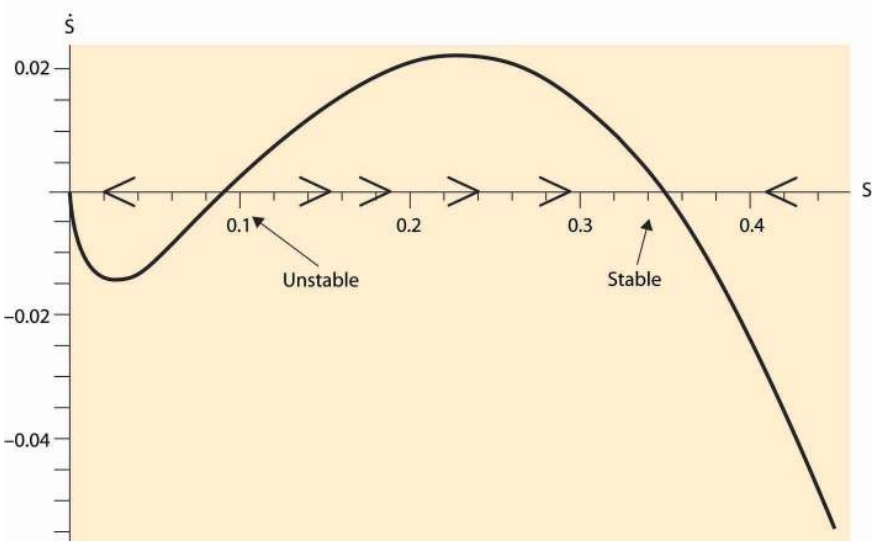


Figure 7.8 Possibility of multiple equilibria

In addition to the stable equilibrium outcome, there is an unstable steady state, which may converge either upward or downward. A feature of fishing with inelastic demand is that there exists a region where extinction is inevitable because, when the stock is near zero, the high demand price induced by inelasticity forces sufficient fishing to ensure extinction.

As a consequence of the fishing externality, nations attempt to regulate fishing, both by extending their own reach 200 miles into the sea and by treaties limiting fishing in the open sea. These regulatory attempts have met with only modest success at preventing overfishing.

What is the efficient stock of fish? This is a challenging mathematical problem, but some insight can be gleaned via a steady-state analysis. A steady state arises when $S^{\cdot} = 0$. If a constant amount Q is removed, a steady state in the stock must occur at $0 = S^{\cdot} = rS(1 - S) - Q$. This maximum catch then occurs at S = ½ and Q = ¼ r. This is not the efficient level, for it neglects the cost of boats, and the efficient stock will actually be larger. More generally, it is never efficient to send the population below the maximum point on the survival curve plotted in Figure 7.5.

Conceptually, fishing is an example of the tragedy of the commons externality already discussed. However, the threat of a permanent extinction and alluring possibility of solving dynamic models make it a particularly dramatic example.

Key Takeaways

- Extinction arises from the interaction of two systems: one biological and one economic.
- When demand is elastic, extinction should not arise.
- When demand is inelastic, population decreases reduce the quantity, which increase total revenue, which leads to more investment in fishing. When demand is sufficiently inelastic, the heightened investment leads to proportionally more fish caught and the fish go extinct.
- Fishing is an example of the tragedy of the commons externality.

EXERCISE

1. Suppose ε = 1. For what parameter values are fish necessarily driven to extinction? Can you interpret this condition to say that the demand for caught fish exceeds the production via reproduction?

This page titled 7.6: Fishing and Extinction is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

8: Public Goods

8.1: Free Riders

8.2: Provision With Taxation

8.3: Local Public Goods

This page titled 8: Public Goods is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

8.1: Free Riders 8.1: Free RidersEdit Update titlecancel

Consider a company offering a fireworks display. Pretty much anyone nearby can watch the fireworks, and people with houses in the right place have a great view of them. The company that creates the fireworks can't compel those with nearby homes to pay for the fireworks, and so a lot of people get to watch them without paying. This will make it difficult or impossible for the fireworks company to make a profit.

Learning Objectives

- What are people who just use public goods without paying called, and what is their effect on economic performance?

A public good is a good that has two attributes: Nonexcludability, which means the producer can't prevent the use of the good by others, and nonrivalry, which means that many people can use the good simultaneously. The classic example of a public good is national defense. National defense is clearly nonexcludable because, if we spend the resources necessary to defend our national borders, it isn't going to be possible to defend everything except one apartment on the second floor of a three-story apartment building on East Maple Street. Once we have kept our enemies out of our borders, we've protected everyone within the borders. Similarly, the defense of the national borders exhibits a fair degree of nonrivalry, especially insofar as the strategy of defense is to deter an attack in the first place. That is, the same expenditure of resources protects all. It is theoretically possible to exclude some people from the use of a poem or a mathematical theorem, but exclusion is generally quite difficult. Both poems and theorems are nonrivalrous. Similarly, technological and software inventions are nonrivalrous, even though a patent grants the right to exclude the use by others. Another good that permits exclusion at a cost is a highway. A toll highway shows that exclusion is possible on the highways. Exclusion is quite expensive, partly because the tollbooths require staffing, but mainly because of the delays imposed on drivers associated with paying the tolls—the time costs of toll roads are high. Highways are an intermediate case where exclusion is possible only at a significant cost, and thus should be avoided if possible. Highways are also rivalrous at high-congestion levels, but nonrivalrous at low-congestion levels. That is, the marginal cost of an additional user is essentially zero for a sizeable number of users, but then marginal cost grows rapidly in the number of users. With fewer than 700 cars per lane per hour on a four-lane highway, generally the flow of traffic is unimpeded.The effect of doubling the number of lanes from two to four is dramatic. A two-lane highway generally flows at 60 mph or more provided there are fewer than 200 cars per lane per hour, while a four-lane highway can accommodate 700 cars per lane per hour at the same speed. As congestion grows beyond this level, traffic slows down and congestion sets in. Thus, west Texas interstate highways are usually nonrivalrous, while Los Angeles's freeways are usually very rivalrous.

Like highways, recreational parks are nonrivalrous at low-use levels, becoming rivalrous as they become sufficiently crowded. Also like highways, it is possible, but expensive, to exclude potential users, since exclusion requires fences and a means for admitting some but not others. (Some exclusive parks provide keys to legitimate users, while others use gatekeepers to charge admission.)

Take the example of a neighborhood association that is considering buying land and building a park in the neighborhood. The value of the park is going to depend on the size of the park, and we suppose for simplicity that the value in dollars of the park to each household in the neighborhood is S b n −a , where n is the number of park users, S is the size of the park, and a and b are parameters satisfying $0 < a \leq b < 1$. This functional form builds in the property that larger parks provide more value at a diminishing rate, but there is an effect from congestion. The functional form gives a reason for parks to be public—it is more efficient for a group of people to share a large park than for each individual to possess a small park, at least if $b > a$, because the gains from a large park exceed the congestion effects. That is, there is a scale advantage—doubling the number of people and the size of the park increases each individual's enjoyment.

How much will selfish individuals voluntarily contribute to the building of the park? That of course depends on what they think others will contribute. Consider a single household, and suppose that each household, i, thinks the others will contribute S_{-1} to the building of the park. Given this expectation, how much should each household, i, contribute? If the household contributes s, the park will have size $S = S_{-1} + s$, which the household values at $(S - 1 + s)bn - a$. Thus, the net gain to a household that contributes s when the others contribute $S_{-1}$is $(S - 1 + s)$ b $n - a - s$.

Exercise 1 shows that individual residents gain from their marginal contribution if and only if the park is smaller than

$$S0 = (bn - a)11 - b$$

. Consequently, under voluntary contributions, the only equilibrium park size is S_0. That is, for any park size smaller than S_0, citizens will voluntarily contribute to make the park larger. For any larger size, no one is willing to contribute.

Under voluntary contributions, as the neighborhood grows in number, the size of the park shrinks. This makes sense—the benefits of individual contributions to the park mostly accrue to others, which reduces the payoff to any one contributor.

How large should the park be? The total value of the park of size S to the residents together is n times the individual value, which gives a collective value of S b n 1−a ; and the park costs S, so from a social perspective the park should be sized to maximize S b n 1−a −S, which yields an optimal park of size $S* = (bn1 - a)11 - b$. Thus, as the neighborhood grows, the park should grow, but as we saw, the park would shrink if the neighborhood has to rely on voluntary contributions. This is because people contribute individually, as if they were building the park for themselves, and don't account for the value they provide to their neighbors when they contribute. Under individual contributions, the hope that others contribute leads individuals not to contribute. Moreover, use of the park by others reduces the value of the park to each individual, so that the size of the park shrinks as the population grows under individual contributions. In contrast, the park ought to grow faster than the number of residents grows, as the per capita park size is $S * n = b11 - bnb - a1 - b$, which is an increasing function of n.Reminder: In making statements like should and ought, there is no conflict in this model because every household agrees about the optimal size of the park, so that a change to a park size of S*, paid with equal contributions, maximizes every household's utility.

The lack of incentive for individuals to contribute to a social good is known as a free-rider problem. The term refers to the individuals who don't contribute to the provision of a public good, who are said to be free riders, that is, they ride freely on the contributions of others. There are two aspects of the free-rider problem apparent in this simple mathematical model. First, the individual incentive to contribute to a public good is reduced by the contributions of others, and thus individual contributions tend to be smaller when the group is larger. Put another way, the size of the free-rider problem grows as the community grows larger. Second, as the community grows larger, the optimal size of the public good grows. The market failure under voluntary contributions is greater as the community is larger. In the theory presented, the optimal size of the public good is $S* = (bn1 - a)11 - b$, and the actual size under voluntary contributions is $S* = (bn1 - a)11 - b$, a gap that gets very large as the number of people grows.

The upshot is that people will voluntarily contribute too little from a social perspective, by free riding on the contributions of others. A good example of the provision of public goods is a coauthored term paper. This is a public good because the grade given to the paper is the same for each author, and the quality of the paper depends on the sum of the efforts of the individual authors. Generally, with two authors, both work pretty hard on the manuscript in order to get a good grade. Add a third author, and it is a virtual certainty that two of the authors will think the third didn't work as hard and is a free rider on the project.

The term paper example also points to the limitations of the theory. Many people are not as selfish as the theory assumes and will contribute more than would be privately optimal. Moreover, with small numbers, bargaining between the contributors and the division of labor (each works on a section) may help to reduce the free-rider problem. Nevertheless, even with these limitations, the free-rider problem is very real; and it gets worse the more people are involved. The theory shows that if some individuals contribute more than their share in an altruistic way, the more selfish individuals contribute even less, undoing some of the good done by the altruists.

Key Takeaways

- A public good has two attributes: nonexcludability, which means the producer can't prevent the use of the good by others; and nonrivalry, which means that many people can use the good simultaneously.
- Examples of public goods include national defense, fireworks displays, and mathematical theorems.
- Nonexcludability implies that people don't have to pay for the good; nonrivalry means that the efficient price is zero.
- A free rider is someone who doesn't pay for a public good.
- Generally voluntary contributions lead to too little provision of public goods.
- In spite of some altruism, the free-rider problem is very real, and it gets worse the more people are involved.

EXERCISES

1. Verify that individual residents gain from contributing to the park if $S < (bn - a)11 - b$ and gain from reducing their contributions if $S > (bn - a)11 - b$..
2. For the model presented in this section, compute the elasticity of the optimal park size with respect to the number of residents—that is, the percentage change in S* for a small percentage change in n. [Hint: Use the linear approximation trick $(1 + \Delta)r \approx r\Delta$

for Δ near zero.]

3. For the model presented in this section, show that an individual's utility when the park is optimally sized and the expenses are shared equally among the n individuals is

$$u = (bb1 - b - b11 - b)nb - a1 - b.$$

   Does this model predict an increase in utility from larger communities?

4. Suppose two people, Person 1 and Person 2, want to produce a playground to share between them. The value of the playground of size S to each person is S , where S is the number of dollars spent to build it. Show that, under voluntary contributions, the size of the playground is ¼ and that the efficient size is 1.
5. For the previous exercise, now suppose Person 1 offers "matching funds"—that is, offers to contribute an equal amount to the contributions of Person 2. How large a playground will Person 2 choose?

This page titled 8.1: Free Riders is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

8.2: Provision With Taxation 8.2: Provision With TaxationEdit Update titlecancel

Learning Objectives

- If people won't pay for public goods, can society tax them instead?

Faced with the fact that voluntary contributions produce an inadequate park, the neighborhood turns to taxes. Many neighborhood associations or condominium associations have taxing authority and can compel individuals to contribute. One solution is to require each resident to contribute the amount 1, resulting in a park that is optimally sized at *n*, as clearly shown in the example from the previous section. Generally it is possible to provide the correct size of the public good using taxes to fund it. However, this is challenging in practice, as we illustrate in this slight modification of the previous example.

Let individuals have different strengths of preferences, so that individual *i* values the public good of size *S* at an amount v i S b n −a that is expressed in dollars. (It is useful to assume that all people have different *v* values to simplify arguments.) The optimal size of the park for the neighborhood is $n - a1 - b(b\sum i = 1nvi)11 - b = (bv^{-})11 - bn1 - a1 - b$, where $v^{-} = 1n\sum i = 1nv$ i is the average value. Again, taxes can be assessed to pay for an optimally sized park, but some people (those with small *v* values) will view that as a bad deal, while others (with large *v*) will view it as a good deal. What will the neighborhood choose to do?

If there are an odd number of voters in the neighborhood, we predict that the park size will appeal most to the **median voter**.The voting model employed here is that there is a status quo, which is a planned size of *S*. Anyone can propose to change the size of *S*, and the neighborhood then votes yes or no. If an *S* exists such that no replacement gets a majority vote, that *S* is an equilibrium under majority voting. This is the voter whose preferences fall in the middle of the range. With equal taxes, an individual obtains v i S b n −a − S n . If there are an odd number of people, *n* can be written as 2*k* + 1. The median voter is the person for whom there are *k* values *vi* larger than hers and *k* values smaller than hers. Consider increasing *S*. If the median voter likes it, then so do all the people with higher *v*'s, and the proposition to increase *S* passes. Similarly, a proposal to decrease *S* will get a majority if the median voter likes it. If the median voter likes reducing *S*, all the individuals with smaller *vi* will vote for it as well. Thus, we can see the preferences of the median voter are maximized by the vote, and simple calculus shows that this entails $S = (bvk)11 - bn1 - a1 - b$.

Unfortunately, voting does not result in an efficient outcome generally and only does so when the average value equals the median value. On the other hand, voting generally performs much better than voluntary contributions. The park size can either be larger or smaller under median voting than is efficient.The general principle here is that the median voting will do better when the distribution of values is such that the average of *n* values exceeds the median, which in turn exceeds the maximum divided by *n*. This is true for most empirically relevant distributions.

Key Takeaways

- Taxation—forced contribution—is a solution to the free-rider problem.
- An optimal tax rate is the average marginal value of the public good.
- Voting leads to a tax rate equal to the median marginal value, and hence does not generally lead to efficiency, although it outperforms voluntary contributions.

EXERCISES

1. Show for the model of this section that, under voluntary contributions, only one person contributes, and that person is the person with the largest *vi*. How much do they contribute? [Hint: Which individual *i* is willing to contribute at the largest park size? Given the park that this individual desires, can anyone else benefit from contributing at all?]
2. Show that, if all individuals value the public good equally, voting on the size of the good results in the efficient provision of the public good.

This page titled 8.2: Provision With Taxation is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

8.3: Local Public Goods 8.3: Local Public GoodsEdit Update titlecancel

Learning Objectives

- What can we do if we disagree about the optimal level of public goods?

The example in the previous section showed the challenges to a neighborhood's provision of public goods created by differences in the preferences. Voting does not generally lead to the efficient provision of the public good and does so only rarely when all individuals have the same preferences.

A different solution was proposed by TieboutCharles Tiebout, 1919–1962. His surname is pronounced "tee-boo." in 1956, which works only when the public goods are local. People living nearby may or may not be excludable, but people living farther away can be excluded. Such goods that are produced and consumed in a limited geographical area are **local public goods**. Schools are local —more distant people can readily be excluded. With parks it is more difficult to exclude people from using the good; nonetheless, they are still local public goods because few people will drive 30 miles to use a park.

Suppose that there are a variety of neighborhoods, some with high taxes, better schools, big parks, beautifully maintained trees on the streets, frequent garbage pickup, a first-rate fire department, extensive police protection, and spectacular fireworks displays, and others with lower taxes and more modest provision of public goods. People will move to the neighborhood that fits their preferences. As a result, neighborhoods will evolve with inhabitants that have similar preferences for public goods. Similarity among neighbors makes voting more efficient, in turn. Consequently, the ability of people to choose their neighborhoods to suit their preferences over taxes and public goods will make the neighborhood provision of public goods more efficient. The "Tiebout theory" shows that local public goods tend to be efficiently provided. In addition, even private goods such as garbage collection and schools can be efficiently publicly provided when they are local goods, and there are enough distinct localities to offer a broad range of services.

Key Takeaways

- When public goods are local—people living nearby may or may not be excludable, whereas people living farther away may be excluded—the goods are "local public goods."
- Specialization of neighborhoods providing in distinct levels of public goods, when combined with households selecting their preferred neighborhood, can lead to efficient provision of public goods.

EXERCISES

1. Consider a babysitting cooperative, where parents rotate supervision of the children of several families. Suppose that, if the sitting service is available with frequency *Y*, a person's *i* value is viY and the costs of contribution $y = (1/2)\text{ny}^2$, where *y* is the sum of the individual contributions and *n* is the number of families. Rank $v_1 \geq v_2 \geq \ldots \geq v_n$.
2. What is the size of the service under voluntary contributions?

   (Hint: Let yi be the contribution of family *i*. Identify the payoff of family *j* as $vj(yj + \sum i \neq jyi) - ½n(yj)2$.

   What value of yjmaximizes this expression?)
3. What contributions maximize the total social value
4. $(\sum j = 1nvj)(\sum j = 1nyj) - ½n\sum i = 1n(yj)2$?
5. yii
6. Let $\mu = 1n\sum j = 1nvj$ and $\sigma 2 = 1n\sum j = 1n(vj - \mu)2$. Conclude that, under voluntary contributions, the total value generated by the cooperative is $n2(\mu 2 - \sigma 2)$.

   (Hint: It helps to know that $\sigma 2 = 1n\sum j = 1n(vj - \mu)2 = 1n\sum j = 1nvj2 - 2n\sum j = 1n\mu vj + 1n\sum j = 1n\mu 2 = 1n\sum j = 1nvj2 - \mu 2$.)

This page titled 8.3: Local Public Goods is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

9: Producer Theory- Costs

9.1: Types of Firms
9.2: Production Functions
9.3: Profit Maximization
9.4: The Shadow Value
9.5: Input Demand
9.6: Myriad Costs

This page titled 9: Producer Theory- Costs is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

9.1: Types of Firms 9.1: Types of FirmsEdit Update titlecancel

The basic theory of the firm regards the firm as a mechanism for transforming productive inputs into final and intermediate goods and services. This process is known as production. For instance, the smelting of copper or gold removes impurities and makes the resulting product into a more valuable substance. Silicon Valley transforms silicon, along with a thousand other chemicals and metals, into computer chips used in everything from computers to toasters. Cooking transforms raw ingredients into food, by adding flavor and killing bacteria. Moving materials to locations where they have higher value is a form of production. Moving stone to a building site to construct an exterior wall, bringing the King Tut museum exhibit to Chicago, or qualifying a basketball team for the league playoffs are all examples of different types of production. According to this simple view, a firm is comprised of a technology or set of technologies for transforming materials into valuable goods and to maximize profits. This "production function" view of the firm is appropriate for some environments, when products and services are standardized and technologies are widely available. However, in other settings, this way of thinking about the firm is misleading, especially when the internal organization of the firm is important. Nevertheless, the "production function" model of the firm is a natural starting point to begin our investigation of competition.

Learning Objectives

- What types of companies exist?

There are four varieties of firms created in law, although these types have several subtypes. At one end is the **proprietorship**, which is a firm owned by a single individual (the proprietor) or perhaps by a family. The family farm and many "mom-and-pop" restaurants and convenience stores are operated proprietorships. Debts accrued by the proprietorship are the personal responsibility of the proprietor. Legal and accounting firms are often organized as **partnerships**. Partnerships are firms owned by several individuals who share profits as well as liabilities of the firm according to a specified formula that varies by the relative contribution and potential cost of each partner in the firm. Thus, if a partner in a law firm steals a client's money and disappears, the other partners may be responsible for absorbing some portion of the loss. In contrast, a **corporation** is treated legally as a single entity owned by shareholders. Like a person, a corporation can incur debt and is therefore responsible for repayment. This stands in contrast to a partnership where particular individuals may be liable for debts incurred. Hence, when the energy trader company Enron collapsed, the Enron shareholders lost the value of their stock; however, they were not responsible for repaying the debt that the corporation had incurred. Moreover, company executives are also not financially responsible for debts of the corporation, provided they act prudentially. If a meteor strikes a manufacturer and destroys the corporation, the executives are not responsible for the damage or for the loans that may not be repaid as a result. On the other hand, executives are not permitted to break the law, as was the case with corporate officers at Archer Daniels Midland, the large agricultural firm who colluded to fix the price of lysine and were subsequently fined and jailed for their misdeeds, as was the corporation.

Corporations who shield company executives and shareholders from fines and punishments are said to offer "limited liability." So why would anyone in his or her right mind organize a firm as a proprietorship or a partnership? The explanation is that it is costly to incorporate businesses—about $1,000 per year at the time of this writing—and corporations are taxed, so that many small businesses find it less costly to be organized as proprietorships. Moreover, it may not be possible for a family-owned corporation to borrow money to open a restaurant; for example, potential lenders may fear not being repaid in the event of bankruptcy, so they insist that owners accept some personal liability. So why are professional groups organized as partnerships and not as corporations? The short answer is that a large variety of hybrid, organizational forms exist instead. The distinctions are blurred, and organizations like "Chapter S Corporations" and "Limited Liability Partnerships" offer the advantages of partnerships (including avoidance of taxation) and corporations. The disadvantages of these hybrids are the larger legal fees and greater restrictions on ownership and freedom to operate that exist in certain states and regions.

Usually proprietorships are smaller than partnerships, and partnerships are smaller than corporations, though there are some very large partnerships (e.g., the Big Four accounting firms) as well as some tiny corporations. The fourth kind of firm may be of any size. It is distinguished primarily by its source of revenue and not by how it is internally organized. The **nonprofit firm** is prohibited from distributing a profit to its owners. Religious organizations, academic associations, environmental groups, most zoos, industry associations, lobbying groups, many hospitals, credit unions (a type of bank), labor unions, private universities, and charities are all organized as nonprofit corporations. The major advantage of nonprofit firms is their tax-free status. In exchange for avoiding taxes, nonprofits must be engaged in government-approved activities, suggesting that nonprofits operate for the social benefit of some segment of society. So why can't you establish your own nonprofit that operates for your personal benefit in order

to avoid taxes? Generally, you alone isn't enough of a socially worthy purpose to meet the requirements to form a nonprofit.Certainly some of the nonprofit religious organizations created by televangelists suggest that the nonprofit established for the benefit of a single individual isn't too far-fetched. Moreover, you can't establish a nonprofit for a worthy goal and not serve that goal but just pay yourself all the money the corporation raises, because nonprofits are prohibited from overpaying their managers. Overpaying the manager means not serving the worthy corporate goal as well as possible. Finally, commercial activities of nonprofits are taxable. Thus, when the nonprofit zoo sells stuffed animals in the gift shop, generally the zoo collects sales tax and is potentially subject to corporate taxes.

The modern corporation is a surprisingly recent invention. Prior to World War I, companies were typically organized as a pyramid with a president at the top and vice presidents who reported to him at the next lower level, and so on. In a pyramid, there is a well-defined chain of command, and no one is ever below two distinct managers of the same level. The problem with a pyramid is that two retail stores that want to coordinate have to contact their managers, and possibly their manager's managers, and so on up the pyramid until a common manager is reached. There are circumstances where such rigid decision making is unwieldy, and the larger the operation of a corporation, the more unwieldy it gets.

Four companies—Sears, DuPont, General Motors, and Standard Oil of New Jersey (Exxon)—found that the pyramid didn't work well for them. Sears found that its separate businesses of retail stores and mail order required a mix of shared inputs (purchased goods) but distinct marketing and warehousing of these goods. Consequently, retail stores and mail order needed to be separate business units, but the purchasing unit has to service both of them. Similarly, DuPont's military business (e.g., explosives) and consumer chemicals were very different operations serving different customers yet often selling the same products so that again the inputs needed to be centrally produced and to coordinate with two separate corporate divisions. General Motors's many car divisions are "friendly rivals," in which technology and parts are shared across the divisions, but the divisions compete in marketing their cars to consumers. Again, technology can't be housed under just one division, but instead is common to all. Finally, Standard Oil of New Jersey was attempting to create a company that managed oil products from oil exploration all the way through to pumping gasoline into automobile gas tanks. With such varied operations all over the globe, Standard Oil of New Jersey required extensive coordination and found that the old business model needed to be replaced. These four companies independently invented the modern corporation, which is organized into separate business units. These business units run as semiautonomous companies themselves, with one business unit purchasing inputs at a negotiated price from another unit and selling outputs to a third unit. The study of the internal organization of firms and its ramifications for competitiveness is fascinating but beyond the scope of this book.If you want to know more about organization theory, I happily recommend *Competitive Solutions: The Strategist's Toolkit*, by R. Preston McAfee, Princeton: Princeton University Press, 2002.

Key Takeaways

- The most basic theory of the firm regards the firm as a means of transforming materials into other, more intermediate and final goods. This is known as production.
- A proprietorship is a firm owned by an individual (the proprietor) or perhaps a family and operated by a small number of people. The family farm and many restaurants and convenience stores are operated in this way. Debts accrued by the proprietorship are the personal responsibility of the proprietor.
- Attorneys and accountants are often organized as partnerships. Partnerships share profits according to a formula and are usually liable for losses incurred by the partnership.
- A corporation is a legal entity that may incur debt, and has the responsibility for repayment of that debt. The owners or officers of the corporation are not liable for the repayment.
- A large variety of hybrid, organizational forms exist.
- The nonprofit firm is prohibited from distributing a profit to its owners. In exchange for avoiding taxes, nonprofits must be engaged in government-approved activities that operate for the benefit of some segment of society. Commercial activities of nonprofits are taxable.
- Large corporations are often organized into business units that are semiautonomous affiliated companies.

This page titled 9.1: Types of Firms is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

9.2: Production Functions 9.2: Production FunctionsEdit Update titlecancel

Learning Objectives

- How is the output of companies modeled?
- Are there any convenient functional forms?

The firm transforms inputs into outputs. For example, a bakery takes inputs like flour, water, yeast, labor, and heat and makes loaves of bread. An earth moving company combines capital equipment, ranging from shovels to bulldozers with labor in order to digs holes. A computer manufacturer buys parts "off-the-shelf" like disk drives and memory, with cases and keyboards, and combines them with labor to produce computers. Starbucks takes coffee beans, water, some capital equipment, and labor to brew coffee.

Many firms produce several outputs. However, we can view a firm that is producing multiple outputs as employing distinct production processes. Hence, it is useful to begin by considering a firm that produces only one output. We can describe this firm as buying an amount x_1 of the first input, x_2 of the second input, and so on (we'll use xn to denote the last input), and producing a quantity of the output. The production function that describes this process is given by $y = f(x_1, x_2, \ldots, x_n)$.

The **production function** is the mapping from inputs to an output or outputs.

For the most part we will focus on two inputs in this section, although the analyses with more than inputs is "straightforward."

Example: The **Cobb-Douglas production function** is the product of each input, x, raised to a given power. It takes the form $f(x_1, x_2, \ldots, x_n)$= a 0 x 1 a 1 x 2 a 2 … x n a n .

The constants a_1 through an are typically positive numbers less than one. For example, with two goods, capital K and labor L, the Cobb-Douglas function becomes $a_0K^aL^b$. We will use this example frequently. It is illustrated, for $a_0 = 1, a = 1/3, \text{ and } b = 2/3$, in Figure 9.1.

Figure 9.1 Cobb-Douglas isoquants

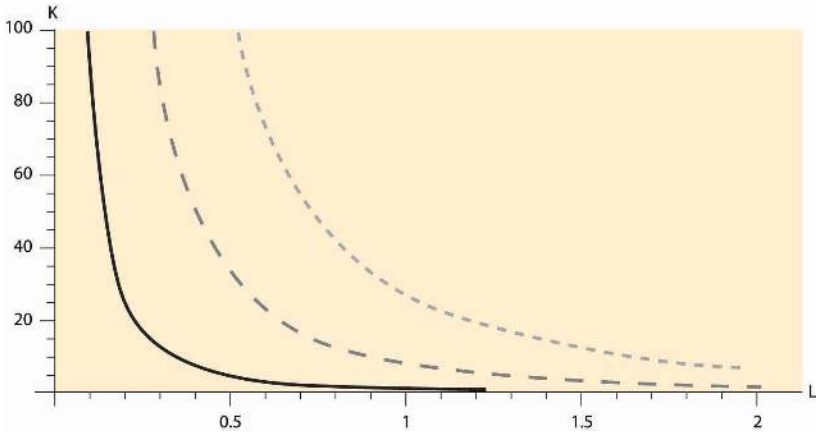


Figure 9.1 The middle, gray dashed line represents an output of 2, and the dotted light-gray line represents an output of 3. Isoquants are familiar contour plots used, for example, to show the height of terrain or temperature on a map. Temperature isoquants are, not surprisingly, called isotherms.

Figure 9.2 The production function

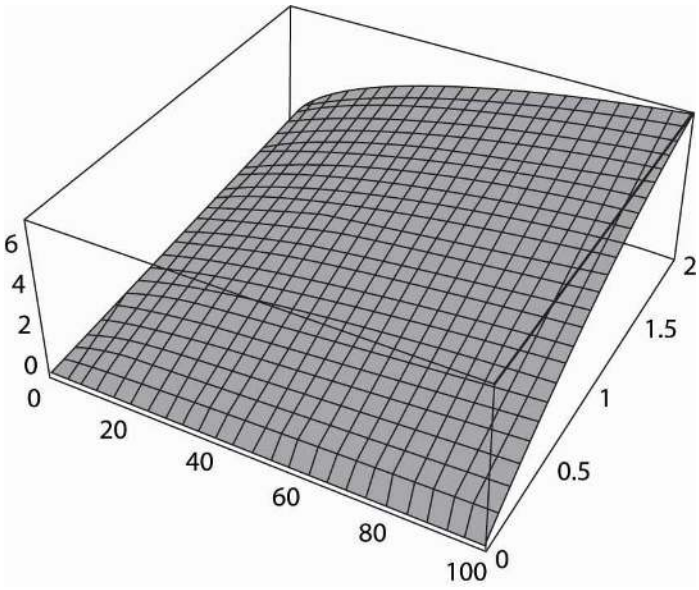


Figure 9.2.

The fixed-proportions production function comes in the form $f(x_1, x_2, \ldots, x_n)$ = Min { a 1 x 1 , a 2 x 2 , …, a n x n }.

The **fixed-proportions production function** is a production function that requires inputs be used in fixed proportions to produce output. It has the property that adding more units of one input in isolation does not necessarily increase the quantity produced. For example, the productive value of having more than one shovel per worker is pretty low, so that shovels and diggers are reasonably modeled as producing holes using a fixed-proportions production function. Moreover, without a shovel or other digging implement like a backhoe, a barehanded worker is able to dig so little that he is virtually useless. Ultimately, the size of the holes is determined by min {number of shovels, number of diggers}. Figure 9.3 illustrates the isoquants for fixed proportions. As we will see, fixed proportions make the inputs “perfect complements.”

Figure 9.3 Fixed-proportions and perfect substitutes

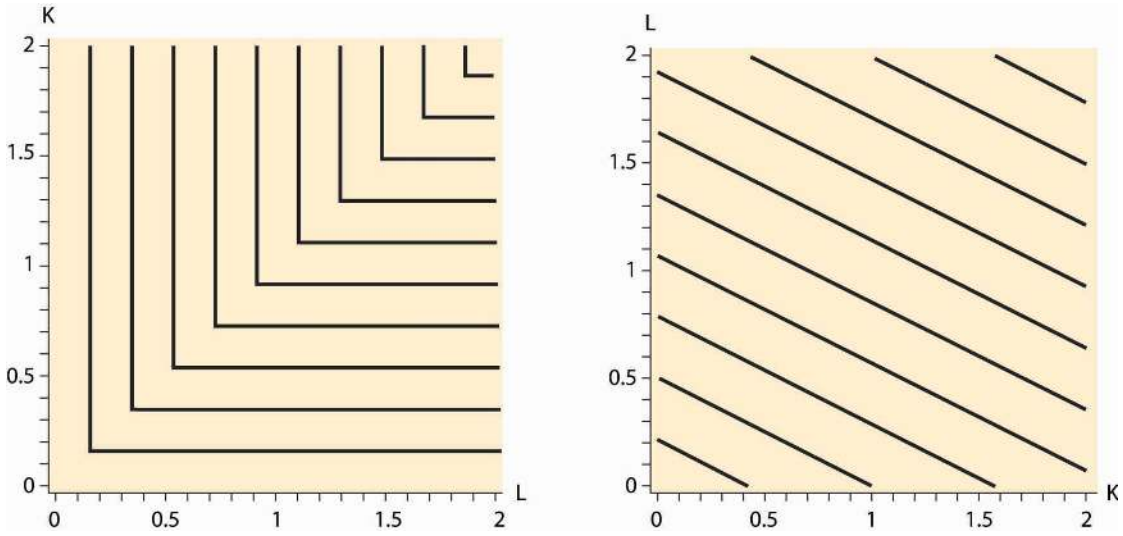


Figure 9.3.

The **marginal product** of an input is just the derivative of the production function with respect to that input.This is a partial derivative, since it holds the other inputs fixed. Partial derivatives are denoted with the symbol δ. An important property of marginal product is that it may be affected by the level of other inputs employed. For example, in the Cobb-Douglas case with two inputsThe symbol α is the Greek letter “alpha.” The symbol β is the Greek letter “beta.” These are the first two letters of the Greek alphabet, and the word alphabet itself originates from these two letters. and for constant A,

$$f(K, L) = AKaL\beta$$

the marginal product of capital is

$$\partial f \partial K(K, L) = \alpha AK\alpha - 1L\beta.$$

If α and β are between zero and one (the usual case), then the marginal product of capital is increasing in the amount of labor, and it is decreasing in the amount of capital employed. For example, an extra computer is very productive when there are many workers and a few computers, but it is not so productive where there are many computers and a few people to operate them.

The **value of the marginal product** of an input is the marginal product times the price of the output. If the value of the marginal product of an input exceeds the cost of that input, it is profitable to use more of the input.

Some inputs are easier to change than others. It can take 5 years or more to obtain new passenger aircraft, and 4 years to build an electricity generation facility or a pulp and paper mill. Very skilled labor such as experienced engineers, animators, and patent attorneys are often hard to find and challenging to hire. It usually requires one to spend 3 to 5 years to hire even a small number of academic economists. On the other hand, it is possible to buy shovels, telephones, and computers or to hire a variety of temporary workers rapidly, in a day or two. Moreover, additional hours of work can be obtained from an existing labor force simply by enlisting them to work "overtime," at least on a temporary basis. The amount of water or electricity that a production facility uses can be varied each second. A dishwasher at a restaurant may easily use extra water one evening to wash dishes if required. An employer who starts the morning with a few workers can obtain additional labor for the evening by paying existing workers overtime for their hours of work. It will likely take a few days or more to hire additional waiters and waitresses, and perhaps several days to hire a skilled chef. You can typically buy more ingredients, plates, and silverware in one day, whereas arranging for a larger space may take a month or longer.

The fact that some inputs can be varied more rapidly than others leads to the notions of the long run and the short run. In the short run, only some inputs can be adjusted, while in the long run all inputs can be adjusted. Traditionally, economists viewed labor as quickly adjustable and capital equipment as more difficult to adjust. That is certainly right for airlines—obtaining new aircraft is a very slow process—for large complex factories, and for relatively low-skilled, and hence substitutable, labor. On the other hand, obtaining workers with unusual skills is a slower process than obtaining warehouse or office space. Generally speaking, the long-run inputs are those that are expensive to adjust quickly, while the short-run factors can be adjusted in a relatively short time frame. What factors belong in which category is dependent on the context or application under consideration.

Key Takeaways

- Firms transform inputs into outputs.
- The functional relationship between inputs and outputs is the production function.
- The Cobb-Douglas production function is the product of the inputs raised to powers and comes in the form $f(x1, x2, \dots, xn) = a0x1a1x2a2 \dots xnan$ for positive constants '_' allowed only in math mode.
- An isoquant is a curve or surface that traces out the inputs leaving the output constant.
- The fixed-proportions production function comes in the form $f(x1, x2, \dots, xn) = mina1x1, a2x2, \dots, anxn$.
- Fixed proportions make the inputs "perfect complements."
- Two inputs K and L are perfect substitutes in a production function f if they enter as a sum; that is, $f(K, L, x_3, \dots, x_n) = g(K + cL, x_3, \dots, x_n)$, for a constant c.
- The marginal product of an input is just the derivative of the production function with respect to that input. An important aspect of marginal products is that they are affected by the level of other inputs.
- The value of the marginal product of an input is just the marginal product times the price of the output. If the value of the marginal product of an input exceeds the cost of that input, it is profitable to use more of the input.
- Some inputs are more readily changed than others.
- In the short run, only some inputs can be adjusted, while in the long run all inputs can be adjusted.
- Traditionally, economists viewed labor as quickly adjustable and capital equipment as more difficult to adjust.
- Generally speaking, the long-run inputs are those that are expensive to adjust quickly, while the short-run factors can be adjusted in a relatively short time frame.

EXERCISE

1. For the Cobb-Douglas production function, suppose there are two inputs K and L, and the sum of the exponents is one. Show that, if each input is paid the value of the marginal product per unit of the input, the entire output is just exhausted. That is, for this production function, show $K\partial f \partial K + L\partial f \partial L = f(K, L)$.

This page titled 9.2: Production Functions is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

9.3: Profit Maximization 9.3: Profit MaximizationEdit Update titlecancel

Learning Objectives

- If firms maximize profits, how will they behave?

Consider an entrepreneur who would like to maximize profit, perhaps by running a delivery service. The entrepreneur uses two inputs, capital K (e.g., trucks) and labor L (e.g., drivers), and rents the capital at cost r per dollar of capital. The wage rate for drivers is w. The production function is F (K, L)—that is, given inputs K and L, the output is F (K, L). Suppose p is the price of the output. This gives a profit of Economists often use the symbol π, the Greek letter "pi," to stand for profit. There is little risk of confusion because economics doesn't use the ratio of the circumference to the diameter of a circle very often. On the other hand, the other two named constants, Euler's e and i, the square root of -1, appear fairly frequently in economic analysis.

$$\pi = pF(K, L) - rK - wL.$$

First, consider the case of a fixed level of K. The entrepreneur chooses L to maximize profit. The value L* of L that maximizes the function π must satisfy

$$0 = \partial\pi\partial L = p\partial F\partial L(K, L*) - w$$

.

This expression is known as a **first-order condition**, a mathematical condition for optimization stating that the first derivative is zero.It is possible that L = 0 is the best that an entrepreneur can do. In this case, the derivative of profit with respect to L is not necessarily zero. The first-order condition instead would be either $0 = \partial\pi\partial L$, or L = 0, and $0 \geq \partial\pi\partial L$. The latter pair of conditions reflects the logic that either the derivative is zero and we are at a maximum, or L = 0, in which case a small increase in L must not cause π to increase. The first-order condition recommends that we add workers to the production process up to the point where the last worker's marginal product is equal to his wage (or cost).

Figure 9.4 Profit-maximizing labor input

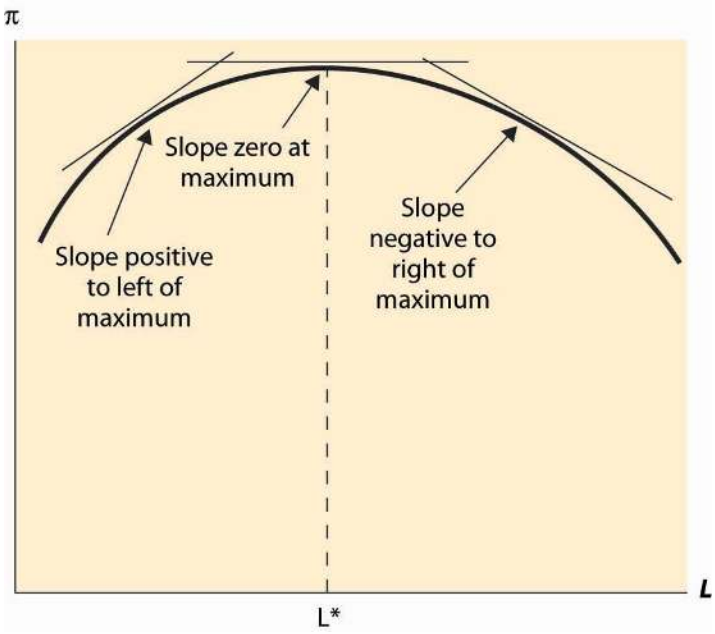


Figure 9.4.

The second property is known as the **second-order condition**, a mathematical condition for maximization stating that the second derivative is nonpositive.The orders refer to considering small, but positive, terms Δ, which are sent to zero to reach derivatives. The value Δ^2, the second-order term, goes to zero faster than Δ, the first-order term. It is expressed as

$$0 \geq \partial 2\pi(\partial L)2 = p\partial 2F(\partial L)2(K, L*).$$

This is enough of a mathematical treatment to establish comparative statics on the demand for labor. Here we treat the choice L* as a function of another parameter—the price p, the wage w, or the level of capital K. For example, to find the effect of the wage on the labor demanded by the entrepreneur, we can write

$$0 = p\partial F\partial L(K, L*(w)) - w.$$

This expression recognizes that the choice L* that the entrepreneur makes satisfies the first-order condition and results in a value that depends on w. But how does it depend on w? We can differentiate this expression to obtain

$$0 = p\partial 2F(\partial L)2(K, L*(w))L*'(w) - 1,$$

or

$$L*'(w) = 1p\partial 2F(\partial L)2(K, L*(w)) \leq 0.$$

The second-order condition enables one to sign the derivative. This form of argument assumes that the choice L* is differentiable, which is not necessarily true.

Digression: In fact, there is a form of argument that makes the point without calculus and makes it substantially more general. Suppose $w_1 < w_2$ are two wage levels and that the entrepreneur chooses L_1 when the wage is w_1 and L_2 when the wage is w_2. Then profit maximization requires that these choices are optimal. In particular, when the wage is w_1, the entrepreneur earns higher profit with L_1 than with L_2:

$$\mathrm{pf(K, L1) - rK - w1L1 \geq pf(K, L2) - rK - w1L2}$$

When the wage is w_2, the entrepreneur earns higher profit with L_2 than with L_1:

$$\mathrm{pf(K, L2) - rK - w2L2 \geq pf(K, L1) - rK - w2L1}$$

.

The sum of the left-hand sides of these two expressions is at least as large as the sum of the right-hand side of the two expressions:

$$\mathrm{pf(K, L1) - rK - w1L1 + pf(K, L2) - rK - w2L2 \geq pf(K, L1) - rK - w2L1 + pf(K, L2) - rK - w1L2}$$

A large number of terms cancel to yield the following:

$$-w1L1 - w2L2 \geq -w2L1 - w1L2$$

This expression can be rearranged to yield the following:

$$(w1 - w2)(L2 - L1) \geq 0$$

This shows that the higher labor choice must be associated with the lower wage. This kind of argument, sometimes known as a **revealed preference** kind of argument, states that choice implies preference. It is called "revealed preference" because choices by consumers were the first place the type of argument was applied. It can be very powerful and general, because issues of differentiability are avoided. However, we will use the more standard differentiability-type argument, because such arguments are usually more readily constructed.

The effect of an increase in the capital level K on the choice by the entrepreneur can be calculated by considering L* as a function of the capital level K:

$$0 = p\partial F\partial L\left(K, L^*(K)\right) - w$$

Differentiating this expression with respect to K, we obtain

$$0 = p\partial 2F\partial K\partial L\left(K, L^*(K)\right) + p\partial 2F(\partial L)2\left(K, L^*(K)\right)L^*(K)$$

or

$$L^{*'}(K) = -\partial 2F\partial K\partial L\left(K, L^*(K)\right)\partial 2F(\partial L)2\left(K, L^*(K)\right)$$

We know the denominator of this expression is not positive, thanks to the second-order condition, so the unknown part is the numerator. We then obtain the conclusion that

an increase in capital increases the labor demanded by the entrepreneur if $\partial 2F\partial K\partial L(K, L*(K)) > 0$, and decreases the labor demanded if $\partial 2\mathrm{F}\partial \mathrm{K}\partial \mathrm{L}\left(\mathrm{K}, \mathrm{L}^{*}(\mathrm{K})\right) < 0$.

This conclusion looks like gobbledygook but is actually quite intuitive. Note that $\partial 2\mathrm{F}\partial \mathrm{K}\partial \mathrm{L}\left(\mathrm{K}, \mathrm{L}^{*}(\mathrm{K})\right) > 0$ means that an increase in capital increases the derivative of output with respect to labor; that is, an increase in capital increases the marginal product of labor. But this is, in fact, the definition of a complement! That is, $\partial 2\mathrm{F}\partial \mathrm{K}\partial \mathrm{L}\left(\mathrm{K}, \mathrm{L}^{*}(\mathrm{K})\right) > 0$ means that labor and capital are complements in production—an increase in capital increases the marginal productivity of labor. Thus, an increase in capital will increase the demand for labor when labor and capital are complements, and it will decrease the demand for labor when labor and capital are substitutes.

This is an important conclusion because different kinds of capital may be complements or substitutes for labor. Are computers complements or substitutes for labor? Some economists consider that computers are complements to highly skilled workers, increasing the marginal value of the most skilled, but substitutes for lower-skilled workers. In academia, the ratio of secretaries to professors has fallen dramatically since the 1970s as more and more professors are using machines to perform secretarial functions. Computers have increased the marginal product of professors and reduced the marginal product of secretaries, so the number of professors rose and the number of secretaries fell.

The revealed preference version of the effect of an increase in capital is to posit two capital levels, K_1 and K_2, with associated profit-maximizing choices L_1 and L_2. The choices require, for profit maximization, that

$$\mathrm{pF(K1, L1) - rK1 - wL1 \geq pF(K1, L2) - rK1 - wL2}$$

and

$$\mathrm{pF(K2, L2) - rK2 - wL2 \geq pF(K2, L1) - rK2 - wL1}$$

Again, adding the left-hand sides together produces a result at least as large as the sum of the right-hand sides:

$$\mathrm{pf}(K1, L1) - KK - wL1 + pF(K2, L2) - rK2 - wL2 \geq pF(K2, L1) - rK2 - wL1 + pF(K1, L2) - rK1 - wL2$$

Eliminating redundant terms yields

$$\mathrm{pF(K1, L1) + pF(K2, L2) \geq pF(K2, L1) + pF(K1, L2)}$$

or

$$\mathrm{F(K2, L2) - F(K1, L2) \geq F(K2, L1) - F(K1, L1)}$$

or

$\int K1K2\partial F\partial K(x, L2)dx \geq \int K1K2\partial F\partial K(x, L1)dx$,Here we use the standard convention that $\int ab \ldots dx = -\int ba \ldots dx$.

or

$$\int K1K2\partial F\partial K(x, L2) - \partial F\partial K(x, L1)dx \geq 0,$$

and finally

$$\int K1K2 \int L1L2\partial 2F\partial K\partial L(x, y)dydx \geq 0.$$

Thus, if $\mathrm{K}_2 > \mathrm{K}_1$ and $\partial 2F\partial K\partial L(K, L) > 0$ for all K and L, then $\mathrm{L}_2 \geq \mathrm{L}_1$; that is, with complementary inputs, an increase in one input increases the optimal choice of the second input. In contrast, with substitutes, an increase in one input decreases the other input. While we still used differentiability of the production function to carry out the revealed preference argument, we did not need to establish that the choice L* was differentiable to perform the analysis.

Example (Labor demand with the Cobb-Douglas production function): The Cobb-Douglas production function has the form $F(K, L) = AK\alpha L\beta$, for constants A, α, and β, all positive. It is necessary for β < 1 for the solution to be finite and well defined. The demand for labor satisfies

$$0 = p\partial F\partial L(K, L*(K)) - w = p\beta AK\alpha L*\beta - 1 - w,$$

or

$$L^{*} = (p\beta AKaw)11 - \beta$$

When $\alpha + \beta = 1$, L is linear in capital. Cobb-Douglas production is necessarily complementary; that is, an increase in capital increases labor demanded by the entrepreneur.

Key Takeaways

- Profit maximization arises when the derivative of the profit function with respect to an input is zero. This property is known as a first-order condition.
- Profit maximization arises with regards to an input when the value of the marginal product is equal to the input cost.
- A second characteristic of a maximum is that the second derivative is negative (or nonpositive). This property is known as the second-order condition.
- Differentiating the first-order condition permits one to calculate the effect of a change in the wage on the amount of labor hired.
- Revealed preference arguments permit one to calculate comparative statics without using calculus, under more general assumptions.
- An increase in capital will increase the demand for labor when labor and capital are complements, and it will decrease the demand for labor when labor and capital are substitutes.
- Cobb-Douglas production functions are necessarily complements; hence, any one input increases as the other inputs increase.

EXERCISES

1. For the fixed-proportions production function min {K, L}, find labor demand when capital is fixed at K.
2. The demand for hamburgers has a constant elasticity of 1 of the form \[$x(p) = 8,000p - 1$. Each entrant in this competitive industry has a fixed cost of $2,000 and produces x hamburgers per year, where x is the amount of meat in pounds.
    1. If the price of meat is $2 per pound, what is the long-run supply of hamburgers?
    2. Compute the equilibrium number of firms, the quantity supplied by each firm, and the market price of hamburgers.
    3. Find the short-run industry supply. Does it have constant elasticity?
3. A company that produces software needs two inputs, programmers (x) at a price of p and computers (y) at a price of r. The output is given by $T = 4x^{1/3}y^{1/3}$, measured in pages of code.
    1. What is the marginal cost?
    2. Now suppose each programmer needs two computers to do his job. What ratio of p and r would make this input mix optimal?
4. A toy factory costs $2 million to construct, and the marginal cost of the q^{th} toy is $\max [10, q^2/1,000]$.
    1. What is average total cost?
    2. What is short-run supply?
    3. What is the long-run competitive supply of toys?

This page titled 9.3: Profit Maximization is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

9.4: The Shadow Value 9.4: The Shadow ValueEdit Update titlecancel

Learning Objectives

- If a firm faces constraints on its behavior, how can we measure the costs of these constraints?

When capital *K* can't be adjusted in the short run, it creates a constraint, on the profit available, to the entrepreneur—the desire to change *K* reduces the profit available to the entrepreneur. There is no direct value of capital because capital is fixed. However, that doesn't mean we can't examine its value. The value of capital is called a **shadow value**, which refers to the value associated with a constraint. Shadow value is well-established jargon.

What is the shadow value of capital? Let's return to the constrained, short-run optimization problem. The profit of the entrepreneur is

$$\pi = pF(K, L) - rK - wL.$$

The entrepreneur chooses the value *L** to maximize profit; however, he is constrained in the short run with the level of capital inherited from a past decision. The shadow value of capital is the value of capital to profit, given the optimal decision *L**. Because $0 = \partial\pi\partial L = p\partial F\partial L(K, L*) - w$, the shadow value of capital is $d\pi(K, L*)dK = \partial\pi(K, L*)\partial K = p\partial F\partial K(K, L*) - r.$

Note that this value could be negative if the entrepreneur might like to sell some capital but can't, perhaps because it is installed in the factory.

Every constraint has a shadow value. The term refers to the value of relaxing the constraint. The shadow value is zero when the constraint doesn't bind; for example, the shadow value of capital is zero when it is set at the profit-maximizing level. Technology binds the firm; the shadow value of a superior technology is the increase in profit associated with it. For example, parameterize the production technology by a parameter *a*, so that *aF*(*K*, *L*) is produced. The shadow value of a given level of *a* is, in the short run,

$$d\pi(K, L*)da = \partial\pi(K, L*)\partial a = pF(K, L*).$$

A term is vanishing in the process of establishing the shadow value. The desired value *L** varies with the other parameters like *K* and *a*, but the effect of these parameters on *L** doesn't appear in the expression for the shadow value of the parameter because $0 = \partial\pi\partial L$ at *L**.

Key Takeaways

- When an input is fixed, its marginal value is called a shadow value.
- A shadow value can be negative when an input is fixed at too high a level.
- Every constraint has a shadow value. The term refers to the value of relaxing a constraint. The shadow value is zero when the constraint doesn't bind.
- The effect of a constraint on terms that are optimized may be safely ignored in calculating the shadow value.

This page titled 9.4: The Shadow Value is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

9.5: Input Demand 9.5: Input DemandEdit Update titlecancel

Learning Objectives

- How much will firms buy?
- How do they respond to input price changes?

Over a long period of time, an entrepreneur can adjust both the capital and the labor used at the plant. This lets the entrepreneur maximize profit with respect to both variables K and L. We'll use a double star, **, to denote variables in their long-run solution. The approach to maximizing profit over two c separately with respect to each variable, thereby obtaining the conditions

$$0 = p\partial F\partial L\left(K^{**}, L^{**}\right) - w$$

and

$$0 = \mathrm{p}\partial \mathrm{F}\partial \mathrm{K}\left(\mathrm{K}^{**}, \mathrm{L}^{**}\right) - \mathrm{r}$$

We see that, for both capital and labor, the value of the marginal product is equal to the purchase price of the input.

It is more challenging to carry out comparative statics exercises with two variables, and the general method won't be developed here.If you want to know more, the approach is to arrange the two equations as a vector with **x** = (K, L), z = (r/p, w/p), so that 0= F ′ (x**)−z, and then differentiate to obtain $dx^{**} = \left(F''\left(x^{**}\right)\right) - 1dz$, which can then be solved for each comparative static. However, we can illustrate one example as follows.

Example: The Cobb-Douglas production function implies choices of capital and labor satisfying the following two first-order conditions:It is necessary for $a + \beta < 1$ for the solution to be finite and well defined.

\begin{equation}0=\mathrm{p} \partial \mathrm{F} \partial \mathrm{L}\left(\mathrm{K}^{* *}, \mathrm{L}^{* *}\right)-\mathrm{w}=\mathrm{p} \beta \mathrm{AK}^{* *} \mathrm{a} \mathrm{L}^{* *} \beta-1-\mathrm{w},

0=\mathrm{p} \partial \mathrm{F} \partial \mathrm{K}\left(\mathrm{K}^{* *}, \mathrm{L}^{* *}\right)-\mathrm{r}=\mathrm{pa} A \mathrm{K}^{* *} \mathrm{a}-1 \mathrm{L}^{* *} \beta-\mathrm{r}\end{equation}

To solve these expressions, first rewrite them to obtain

$$\mathrm{w} = \mathrm{p}\beta \mathrm{AK}^{**}\mathrm{aL}^{**}\beta - 1$$

and

$$r = pqAK^{**}a - 1L^{**}\beta$$

Then divide the first expression by the second expression to yield

$$\mathrm{wr} = \beta \mathrm{K}^{**}\mathrm{al}^{**}$$

or

$$\mathrm{K}^{**} = \mathrm{awBrL}^{**}$$

This can be substituted into either equation to obtain

$$L^{**} = (\ \mathrm{Ap\ a\ a}\ \beta 1 - a\ \mathrm{r\ a}\ w1 - a\ )\ 11 - a - \beta$$

and

$$\mathrm{K}^{**} = (\mathrm{Apa}1 - \beta\beta\beta \mathrm{r}1 - \beta \mathrm{w}\beta)11 - \mathrm{a} - \beta$$

.

While these expressions appear complicated, the dependence on the output price p, and the input prices r and w, is quite straightforward.

How do equilibrium values of capital and labor respond to a change in input prices or output price for the Cobb-Douglas production function? It is useful to cast these changes in percentage terms. It is straightforward to demonstrate that both capital and labor respond to a small percentage change in any of these variables with a constant percentage change.

An important insight of profit maximization is that it implies minimization of costs of yielding the chosen output; that is, profit maximization entails efficient production.

The logic is straightforward. The profit of an entrepreneur is revenue minus costs, and the revenue is price times output. For the chosen output, then, the entrepreneur earns the revenue associated with the output, which is fixed since we are considering only the chosen output, minus the costs of producing that output. Thus, for the given output, maximizing profits is equivalent to maximizing a constant (revenue) minus costs. Since maximizing –C is equivalent to minimizing C, the profit-maximizing entrepreneur minimizes costs. This is important because profit-maximization implies not being wasteful in this regard: A profit-maximizing entrepreneur produces at least cost.

Figure 9.5 Tangency and Isoquants

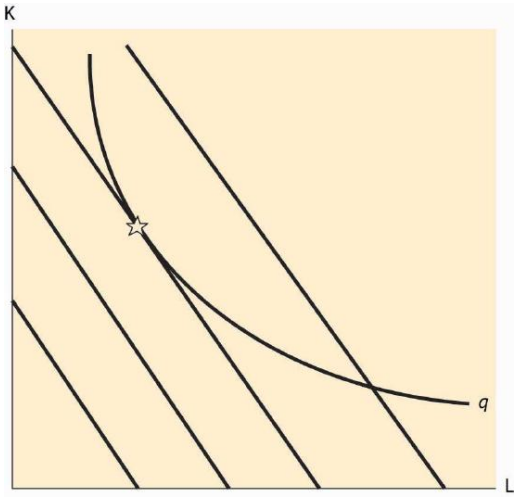


Figure 9.5. The curve represents an isoquant, which holds constant the output. The straight lines represent **isocost** lines, which hold constant the expenditure on inputs. Isocost lines solve the problem $rK + wL = constant$ and thus have slope $dKdL = -wr$. Isocost lines are necessarily parallel—they have the same slope. Moreover, the cost associated with an isocost line rises the farther northeast we go in the graph, or the farther away from the origin.

What point on an isoquant minimizes total cost? The answer is the point associated with the lowest (most southwest) isocost that intersects the isoquant. This point will be tangent to the isoquant and is denoted by a star. At any lower cost, it isn't possible to produce the desired quantity. At any higher cost, it is possible to lower cost and still produce the quantity.

The fact that cost minimization requires a tangency between the isoquant and the isocost has a useful interpretation. The slope of the isocost is minus the ratio of input prices. The slope of the isoquant measures the substitutability of the inputs in producing the output. Economists call this slope the **marginal rate of technical substitution**, which is the amount of one input needed to make up for a decrease in another input while holding output constant. Thus, one feature of cost minimization is that the input price ratio equals the marginal rate of technical substitution.

Key Takeaways

- In the long run, all inputs can be optimized, which leads to multiple first-order conditions.
- The solution can be illustrated graphically and computed explicitly for Cobb-Douglas production functions.
- An important implication of profit maximization is cost minimization—output is produced by the most efficient means possible.
- Cost minimization occurs where the ratio of the input prices equals the slope of the isocost curve, known as the marginal rate of technical substitution, which is the amount of one input needed to make up for a decrease in another input and hold output constant.

EXERCISE

For the Cobb-Douglas production function $F(K, L) = AK$ a $L\beta$, show that $rL^{**}\partial L^{**}\partial r = -\alpha 1 - \alpha - \beta$, $wL^{**}\partial L^{**}\partial w = -1 - \alpha 1 - \alpha - \beta$, $pL^{**}\partial L^{**}\partial p = 11 - \alpha - \beta, rK^{**}\partial K^{**}\partial r = -1 - \beta 1 - \alpha - \beta$, $wK^{*} * \partial K^{*} * \partial w = -\beta 1 - a - \beta$ and $pK^{**}\partial K^{**}\partial p = 11 - \alpha - \beta$.

This page titled 9.5: Input Demand is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

9.6: Myriad Costs 9.6: Myriad CostsEdit Update titlecancel

Learning Objectives

- What are the different ways of measuring costs, and how are they related to the amount of time the firm has to change its behavior?

How much does it cost to produce a given quantity q? We already have a detailed answer to this question, but now we need to focus less on the details and more on the "big picture." First, let's focus on the short run and suppose L is adjustable in the short run but K is not. Then the **short-run total cost** of producing q,—that is, the total cost of output with only short-term factors varying—given the capital level, is

$$\mathrm{SRTC(q|K) = min\, L\ rK{+}wL,\ over\ all\ L\ satisfying\ F(K,L) \geq q}$$

In words, this equation says that the short-run total cost of the quantity q, given the existing level K, is the minimum cost, where L gets to vary (which is denoted by "min over L") and where the L considered is large enough to produce q. The vertical line | is used to indicate a condition or conditional requirement; here |K indicates that K is fixed. The minimum lets L vary but not K. Finally, there is a constraint $F(K, L) \geq q$, which indicates that one must be able to produce q with the mix of inputs because we are considering the short-run cost of q.

The short-run total cost of q, given K, has a simple form. First, since we are minimizing cost, the constraint F(K, L) ≥ q will be satisfied with equality, F(K, L) = q. This equation determines L, since K is fixed; that is, $\mathrm{F(K, LS(q, K)) = q}$ gives the short-run value of L, LS(q, K). Finally, the cost is then $\mathrm{F(K, LS(q, K)) = q}$.

The **short-run marginal cost**, given K, is just the derivative of short-run total cost with respect to output, q. To establish the short-run marginal cost, note that the equation $F(K, L) = q$ implies

$$\mathrm{aF\ \partial L(K, LS(q, K))dL = dq}$$

or

$$\mathrm{dL\ dq\ |F = q = 1\ dF\ \partial L(K, LS(q, K))}$$

The tall vertical line, subscripted with F = q, is used to denote the constraint F(K, L) = q that is being differentiated. Thus, the short-run marginal cost is

$$\mathrm{SRMC(q|K) = SRT\, C'(q) = ddq(rK + wL) = w\ dL\ dq|F = q = w\partial F\partial L(K, LS(q, K))}$$

There are three other short-run costs we require to complete the analysis. First, there is the short-run average cost of production that we obtain by dividing the total cost by the quantity produced:

$$\mathrm{SRAC}(q|\mathrm{K}) = \mathrm{SRTC(q|K)q}$$

Second, there is the **short-run variable cost** that is the total cost minus the cost of producing zero units—that is, minus the fixed cost—which in this case is rK. Finally, we need one more short-run cost: the short-run average variable cost. The short-run average variable cost is the short-run variable cost divided by quantity, which is given,

$$\mathrm{SRAVC(q|K) = SRT(q|K) - SRTC(0|K)q = wLS(q|K)q}$$

The **short-run average variable cost** is the average cost ignoring the investment in capital equipment.

The **short-run average cost** could also be called the short-run average total cost, since it is the average of the total cost per unit of output, but "average total" is a bit of an oxymoron.An oxymoron is a word or phrase that is self-contradictory, like "jumbo shrimp," "stationary orbit," "virtual reality," "modern tradition," or "pretty ugly." Oxymoron comes from the Greek "oxy," meaning sharp, and "moros," meaning dull. Thus, oxymoron is itself an oxymoron, so an oxymoron is self-descriptive. Another word that is self-descriptive is "pentasyllabic." Consequently, when total, fixed, or variable is not specified, the convention is to mean total. Note that the marginal variable cost is the same as the marginal total costs, because the difference between variable cost and total cost is a constant—the cost of zero production, also known as the **short-run fixed cost** of production.

At this point, we have identified four distinct costs, all of which are relevant to the short run. These are the total cost, the marginal cost, the average cost, and the average variable cost. In addition, all of these costs may be considered in the long run as well. There

are three differences in the long run. First, the long run lets all inputs vary, so the **long-run total cost** is the total cost of output with all factors varying. In this case, $\mathrm{LRTC(q)} = \min \mathrm{L, KrK + wL}$ over all L and K combinations satisfying $F(K, L) \geq q$.

Second, since all inputs can vary, the long-run cost isn't conditioned on K. Finally, the **long-run average variable cost** is the long-run total cost divided by output; it is also known as the long-run average total cost. Since a firm could use no inputs in the long run and thus incur no costs, the cost of producing zero is zero. Therefore, in the long run, all costs are variable, and the long-run average variable cost is the long-run average total cost divided by quantity.

Note that the easiest way to find the long-run total cost is to minimize the short-run total cost over K. Since this is a function of one variable, it is straightforward to identify the K that minimizes cost, and then plug that K into the expression for total cost.

One might want to distinguish the very short run from the short run, the medium run, the long run, and the very long run.. But a better approach is to view adjustment as a continuous process, with a gradual easing of the constraints. Faster adjustment costs more. Continuous adjustment is a more advanced topic, requiring an Euler equation approach.

Key Takeaways

- The short-run total cost is the minimum cost of producing a given quantity minimized over the inputs variable in the short run. Sometimes the word total is omitted.
- The short-run fixed cost is the short-run total cost at a zero quantity.
- The short-run marginal cost, given K, is just the derivative of the short-run total cost with respect to quantity.
- The short-run average cost is the short-run total cost divided by quantity.
- The short-run average variable cost is the short-run total cost minus the short-run fixed cost, all divided by quantity.
- Marginal variable cost is the same as the marginal total costs, because the difference between total cost and variable cost is the fixed cost of production, a constant.
- The long-run total cost is the minimum cost of producing a given quantity minimized over all inputs. Sometimes the word total is omitted.
- The long-run fixed cost is zero.
- The long-run marginal cost is the derivative of the long-run total cost with respect to quantity.
- The long-run average cost is the long-run total cost divided by quantity.
- The long-run average variable cost equals the long-run average cost.

EXERCISES

1. For the Cobb-Douglas production function $\mathrm{F(K, L)} = \mathrm{AKaL}\beta$, with $a + \beta < 1$, K fixed in the short run but not in the long run, and cost r of capital and w for labor, show

$$\mathrm{SRTC(q|K)} = \mathrm{rK} + \mathrm{w(qAKa)1}\beta$$

,

$$\mathrm{SRAVC(q|K)} = \mathrm{wq1} - \beta\beta\mathrm{(AKa)1}\beta$$

,

$$\mathrm{SRMC(q|K)} = \mathrm{w} \quad \mathrm{q1} - \beta\beta\beta\mathrm{(AKa)1}\beta$$

,

$$\angle RTC(q|K) = ((a\beta)\beta a + \beta + (\beta a)aa + \beta)w\beta a + \beta raa + \beta(qA)1a + \beta$$

.

2. Consider a cost function of producing an output q of the form $c(a) = q^2 + 2q + 16$. Determine the following:
    1. Marginal cost
    2. Average cost
    3. Average variable cost

    Graph the long-run supply curve, assuming the cost function is for a single plant and can be replicated without change.

This page titled 9.6: Myriad Costs is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

10: Producer Theory- Dynamics

10.1: Reactions of Competitive Firms
10.2: Economies of Scale and Scope
10.3: Dynamics With Constant Costs
10.4: General Long-run Dynamics

This page titled 10: Producer Theory- Dynamics is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

10.1: Reactions of Competitive Firms 10.1: Reactions of Competitive FirmsEdit Update titlecancel

Learning Objectives

- How does a competitive firm respond to price changes?

In this section, we consider a competitive firm (or entrepreneur) that can't affect the price of output or the prices of inputs. How does such a competitive firm respond to price changes? When the price of the output is p, the firm earns profits $\pi = pq - c(q|K)$, where $c(q|K)$ is the total cost of producing, given that the firm currently has capital K. Assuming that the firm produces at all, it maximizes profits by choosing the quantity qs satisfying $0 = p - c'(qs|K)$; which is the quantity where price equals marginal cost. However, this is a good strategy only if producing a positive quantity is desirable, so that $\mathrm{p}\ qs - c(qs|K) \geq -c(0, K)$, which maybe rewritten as $p \geq c(qs|K) - c(0, K)qs$. The right-hand side of this inequality is the average variable cost of production, and thus the inequality implies that a firm will produce, provided price exceeds the average variable cost. Thus, the profit-maximizing firm produces the quantity qs, where price equals marginal cost, provided price is as large as minimum average variable cost. If price falls below minimum average variable cost, the firm shuts down.

The behavior of the competitive firm is illustrated in Figure 10.1. The thick line represents the choice of the firm as a function of the price, which is on the vertical axis. Thus, if the price is below the minimum average variable cost (AVC), the firm shuts down. When price is above the minimum average variable cost, the marginal cost gives the quantity supplied by the firm. Thus, the choice of the firm is composed of two distinct segments: the marginal cost, where the firm produces the output such that price equals marginal cost; and shutdown, where the firm makes a higher profit, or loses less money, by producing zero.

Figure 10.1 also illustrates the average total cost, which doesn't affect the short-term behavior of the firm but does affect the long-term behavior because, when price is below average total cost, the firm is not making a profit. Instead, it would prefer to exit over the long term. That is, when the price is between the minimum average variable cost and the minimum average total cost, it is better to produce than to shut down; but the return on capital was below the cost of capital. With a price in this intermediate area, a firm would produce but would not replace the capital, and thus would shut down in the long term if the price were expected to persist. As a consequence, minimum average total cost is the long-run "shutdown" point for the competitive firm. (Shutdown may refer to reducing capital rather than literally setting capital to zero.) Similarly, in the long term, the firm produces the quantity where the price equals the long-run marginal cost.

Figure 10.1 Short-run supply

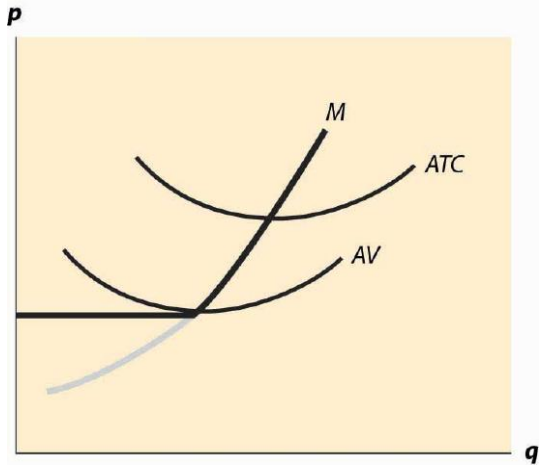


Figure 10.1 illustrates one other fact: The minimum of average cost occurs at the point where marginal cost equals average cost. To see this, let C(q) be total cost, so that average cost is C(q)/q. Then the minimum of average cost occurs at the point satisfying

$$0 = \mathrm{ddq} \quad \mathrm{C(q)q} = \mathrm{C'(q)q} - \mathrm{C(q)q2}$$

But this can be rearranged to imply $C'(q) = C(q)q$, where marginal cost equals average cost at the minimum of average cost.

The long-run marginal cost has a complicated relationship to short-run marginal cost. The problem in characterizing the relationship between long-run and short-run marginal costs is that some costs are marginal in the long run that are fixed in the short

run, tending to make long-run marginal costs larger than short-run marginal costs. However, in the long run, the assets can be configured optimally, While some assets are fixed in the short run, and this optimal configuration tends to make long-run costs lower.

Instead, it is more useful to compare the long-run average total costs and short-run average total costs. The advantage is that capital costs are included in short-run average total costs. The result is a picture like Figure 10.2.

Figure 10.2 Average and marginal costs

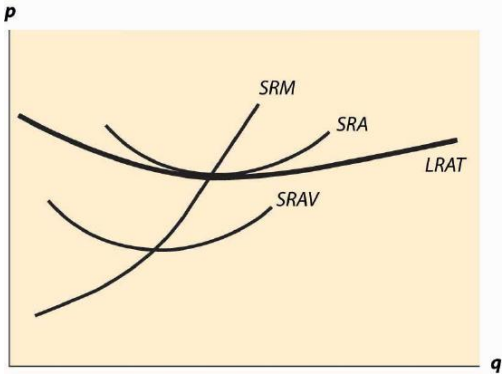


Figure 10.3.

In Figure 10.3, the quantity produced is larger than the quantity that minimizes long-run average total cost. Consequently, as is visible in the figure, the quantity where short-run average cost equals long-run average cost does not minimize short-run average cost. What this means is that a factory designed to minimize the cost of producing a particular quantity won't necessarily minimize short-run average cost. Essentially, because the long-run average total cost is increasing, larger plant sizes are getting increasingly more expensive, and it is cheaper to use a somewhat "too small" plant and more labor than the plant size with the minimum short-run average total cost. However, this situation wouldn't likely persist indefinitely because, as we shall see, competition tends to force price to the minimum long-run average total cost. At this point, then, we have the three-way equality between long-run average total cost, short-run average total cost, and short-run marginal cost.

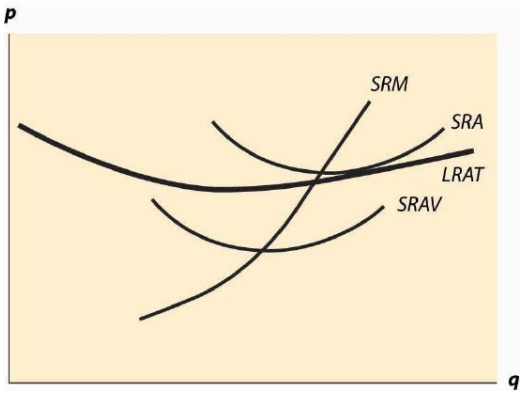


Figure 10.3 Increased plant size

Key Takeaways

- The profit-maximizing firm produces the quantity where price equals marginal cost, provided price is as large as minimum average variable cost. If price falls below minimum average variable cost, the firm shuts down.
- When price falls below short-run average cost, the firm loses money. If price is above average variable cost, the firm loses less money than it would by shutting down; once price falls below short-run average variable cost, shutting down entails smaller losses than continuing to operate.
- The minimum of average cost occurs at the point where marginal cost equals average cost.
- If price is below long-run average cost, the firm exits in the long run.
- Every point on long-run average total cost must be equal to a point on some short-run average total cost.
- The quantity where short-run average cost equals long-run average cost need not minimize short-run average cost if long-run average cost isn't constant.

EXERCISE

1. Suppose a company has total cost given by $rK + q22K$, where capital K is fixed in the short run. What is short-run average total cost and what is marginal cost? Plot these curves. For a given quantity q_0, what level of capital minimizes total cost? What is the minimum average total cost of q_0?

This page titled 10.1: Reactions of Competitive Firms is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

10.2: Economies of Scale and Scope 10.2: Economies of Scale and ScopeEdit Update titlecancel

Learning Objectives

- When firms get bigger, when do average costs rise or fall?
- How does size relate to profit?

An **economy of scale**—that larger scale lowers cost—arises when an increase in output reduces average costs. We met economies of scale and its opposite, diseconomies of scale, in the previous section, with an example where long-run average total cost initially fell and then rose, as quantity was increased.

What makes for an economy of scale? Larger volumes of productions permit the manufacture of more specialized equipment. If I am producing a million identical automotive taillights, I can spend $50,000 on an automated plastic stamping machine and only affect my costs by 5 cents each. In contrast, if I am producing 50,000 units, the stamping machine increases my costs by a dollar each and is much less economical.

Indeed, it is somewhat more of a puzzle to determine what produces a diseconomy of scale. An important source of diseconomies is managerial in nature—organizing a large, complex enterprise is a challenge, and larger organizations tend to devote a larger percentage of their revenues to management of the operation. A bookstore can be run by a couple of individuals who rarely, if ever, engage in management activities, where a giant chain of bookstores needs finance, human resource, risk management, and other "overhead" type expenses just in order to function. Informal operation of small enterprises is replaced by formal procedural rules in large organizations. This idea of managerial diseconomies of scale is reflected in the aphorism "A platypus is a duck designed by a committee."

In his influential 1975 book The Mythical Man-Month, IBM software manager Fred Books describes a particularly severe diseconomy of scale. Adding software engineers to a project increases the number of conversations necessary between pairs of individuals. If there are n engineers, there are $½n(n–1)$ pairs, so that communication costs rise at the square of the project size. This is pithily summarized in Brooks' Law: "Adding manpower to a late software project makes it later."

Another related source of diseconomies of scale involves system slack. In essence, it is easier to hide incompetence and laziness in a large organization than in a small one. There are a lot of familiar examples of this insight, starting with the Peter Principle, which states that people rise in organizations to the point of their own incompetence, meaning that eventually people cease to do the jobs that they do well.Laurence Johnston Peter (1919–1990). The notion that slack grows as an organization grows implies a diseconomy of scale.

Generally, for many types of products, economies of scale from production technology tend to reduce average cost, up to a point where the operation becomes difficult to manage. Here the diseconomies tend to prevent the firm from economically getting larger. Under this view, improvements in information technologies over the past 20 years have permitted firms to get larger and larger. While this seems logical, in fact firms aren't getting that much larger than they used to be; and the share of output produced by the top 1,000 firms has been relatively steady; that is, the growth in the largest firms just mirrors world output growth.

Related to an economy of scale is an **economy of scope**. An economy of scope is a reduction in cost associated with producing several distinct goods. For example, Boeing, which produces both commercial and military jets, can amortize some of its research and development (R&D) costs over both types of aircraft, thereby reducing the average costs of each. Scope economies work like scale economies, except that they account for advantages of producing multiple products, where scale economies involve an advantage of multiple units of the same product.

Economies of scale can operate at the level of the individual firm but can also operate at an industry level. Suppose there is an economy of scale in the production of an input. For example, there is an economy of scale in the production of disk drives for personal computers. This means that an increase in the production of PCs will tend to lower the price of disk drives, reducing the cost of PCs, which is a scale economy. In this case, it doesn't matter to the scale economy whether one firm or many firms are responsible for the increased production. This is known as an **external economy of scale**, or an industry economy of scale, because the scale economy operates at the level of the industry rather than in the individual firm. Thus, the long-run average cost of individual firms may be flat, while the long-run average cost of the industry slopes downward.

Even in the presence of an external economy of scale, there may be diseconomies of scale at the level of the firm. In such a situation, the size of any individual firm is limited by the diseconomy of scale, but nonetheless the average cost of production is decreasing in the total output of the industry, through the entry of additional firms. Generally there is an external diseconomy of scale if a larger industry drives up input prices; for example, increasing land costs. Increasing the production of soybeans significantly requires using land that isn't so well suited for them, tending to increase the average cost of production. Such a diseconomy is an external diseconomy rather than operating at the individual farmer level. Second, there is an external economy if an increase in output permits the creation of more specialized techniques and a greater effort in R&D is made to lower costs. Thus, if an increase in output increases the development of specialized machine tools and other production inputs, an external economy will be present.

An economy of scale arises when total average cost falls as the number of units produced rises. How does this relate to production functions? We let $y = f(x1, x2, \dots, xn)$ be the output when the n inputs $x_1, x_2, \dots, xn$ are used. A rescaling of the inputs involves increasing the inputs by a fixed percentage; e.g., multiplying all of them by the constant λ (the Greek letter "lambda"), where λ > 1. What does this do to output? If output goes up by more than λ, we have an economy of scale (also known as **increasing returns to scale**): Scaling up production increases output proportionately more. If output goes up by less than λ, we have a diseconomy of scale, or **decreasing returns to scale**. And finally, if output rises by exactly λ, we have **constant returns to scale**. How does this relate to average cost? Formally, we have an economy of scale if $f(\lambda x1, \lambda x2, \dots, \lambda xn) > \lambda f(x1, x2, \dots, xn) if \lambda > 1.$

This corresponds to decreasing average cost. Let w_1 be the price of input one, w_2 the price of input two, and so on. Then the average cost of producing $y = f(x1, x2, \dots, xn)$ is $AVC = w1 \times 1 + w2 \times 2 + \dots + wn \times nf(x1, x2, \dots, xn)$

What happens to average cost as we scale up production by λ > 1? Call this AVC(λ).

$$AVC(\lambda) = w1\lambda x1 + w2\lambda x2 + \dots + wn\lambda xnf(\lambda x1, \lambda x2, \dots, \lambda xn) = \lambda w1x1 + w2x2 + \dots + wnxnf(\lambda x1, \lambda x2, \dots, \lambda xn)$$
$$= \lambda f(x1, x2, \dots, xn) f(\lambda x1, \lambda x2, \dots, \lambda xn) AVC(1)$$

Thus, average cost falls if there is an economy of scale and rises if there is a diseconomy of scale.

Another insight about the returns to scale concerns the value of the marginal product of inputs. Note that if there are constant returns to scale, then

$$x1\partial f \partial x1 + x2\partial f \partial x2 + \dots + xn\partial f \partial xn = dd\lambda f(\lambda x1, \lambda x2, \dots, \lambda xn)$$
$$|\lambda \to 1 = lim\lambda \to 1 f(\lambda x1, \lambda x2, \dots, \lambda xn) - f(x1, x2, \dots, xn)\lambda - 1 = f(x1, x2, \dots, xn).$$

The value ∂f ∂ x 1 is the marginal product of input x_1, and similarly ∂f ∂ x 2 is the marginal product of the second input, and so on. Consequently, if the production function exhibits constant returns to scale, it is possible to divide up output in such a way that each input receives the value of the marginal product. That is, we can give x 1 ∂f ∂ x 1 to the suppliers of input one, x 2 ∂f ∂ x 2 to the suppliers of input two, and so on; and this exactly uses up all of the output. This is known as "paying the marginal product," because each supplier is paid the marginal product associated with the input.

If there is a diseconomy of scale, then paying the marginal product is feasible; but there is generally something left over, too. If there are increasing returns to scale (an economy of scale), then it is not possible to pay all the inputs their marginal product; that is, $x1\partial f \partial x1 + x2\partial f \partial x2 + \dots + xn\partial f \partial xn > f(x1, x2, \dots, xn).$

Key Takeaways

- An economy of scale arises when an increase in output reduces average costs.
- Specialization may produce economies of scale.
- An important source of diseconomies is managerial in nature—organizing a large, complex enterprise is a challenge, and larger organizations tend to devote a larger percentage of their revenues to management of the operation.
- An economy of scope is a reduction in cost associated with producing several related goods.
- Economies of scale can operate at the level of the individual firm but can also operate at an industry level. At the industry level, scale economies are known as an external economies of scale or an industry economies of scale.
- The long-run average cost of individual firms may be flat, while the long-run average cost of the industry slopes downward.
- Generally there is an external diseconomy of scale if a larger industry drives up input prices. There is an external economy if an increase in output permits the creation of more specialized techniques and a greater effort in R&D is made to lower costs.
- A production function has increasing returns to scale if an increase in all inputs by a constant factor λ increases output by more than λ.

- A production function has decreasing returns to scale if an increase in all inputs by a constant factor λ increases output by less than λ.
- The production function exhibits increasing returns to scale if and only if the cost function has an economy of scale.
- When there is an economy of scale, the sum of the values of the marginal product exceeds the total output. Consequently, it is not possible to pay all inputs their marginal product.
- When there is a diseconomy of scale, the sum of the values of the marginal product is less than the total output. Consequently, it is possible to pay all inputs their marginal product and have something left over for the entrepreneur.

EXERCISES

1. Given the Cobb-Douglas production function $f(x1, x2, \ldots, xn) = x1$ a 1×2 a $2 \ldots x$ n an , show that there is constant returns to scale if $a1 + a2 + \ldots + an = 1$, increasing returns to scale if $a1 + a2 + \ldots + an > 1$, and decreasing returns to scale if $a1 + a2 + \ldots + an < 1$.
2. Suppose a company has total cost given by $[K + g22K]$, where capital K can be adjusted in the long run. Does this company have an economy of scale, diseconomy of scale, or constant returns to scale in the long run?
3. A production function f is homogeneous of degree r if $f(\lambda x1, \lambda x2, \ldots, \lambda xn) = \lambda r f(x1, x2, \ldots, xn)$. Consider a firm with a production function that is homogeneous of degree r. Suppose further that the firm pays the value of marginal product for all of its inputs. Show that the portion of revenue left over is $1–r$.

This page titled 10.2: Economies of Scale and Scope is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

10.3: Dynamics With Constant Costs 10.3: Dynamics With Constant CostsEdit Update titlecancel

Learning Objectives

- How do changes in demand or cost affect the short- and long-run prices and quantities traded?

Having understood how a competitive firm responds to price and input cost changes, we consider how a competitive market responds to demand or cost changes.

Figure 10.4 Long-run equilibrium

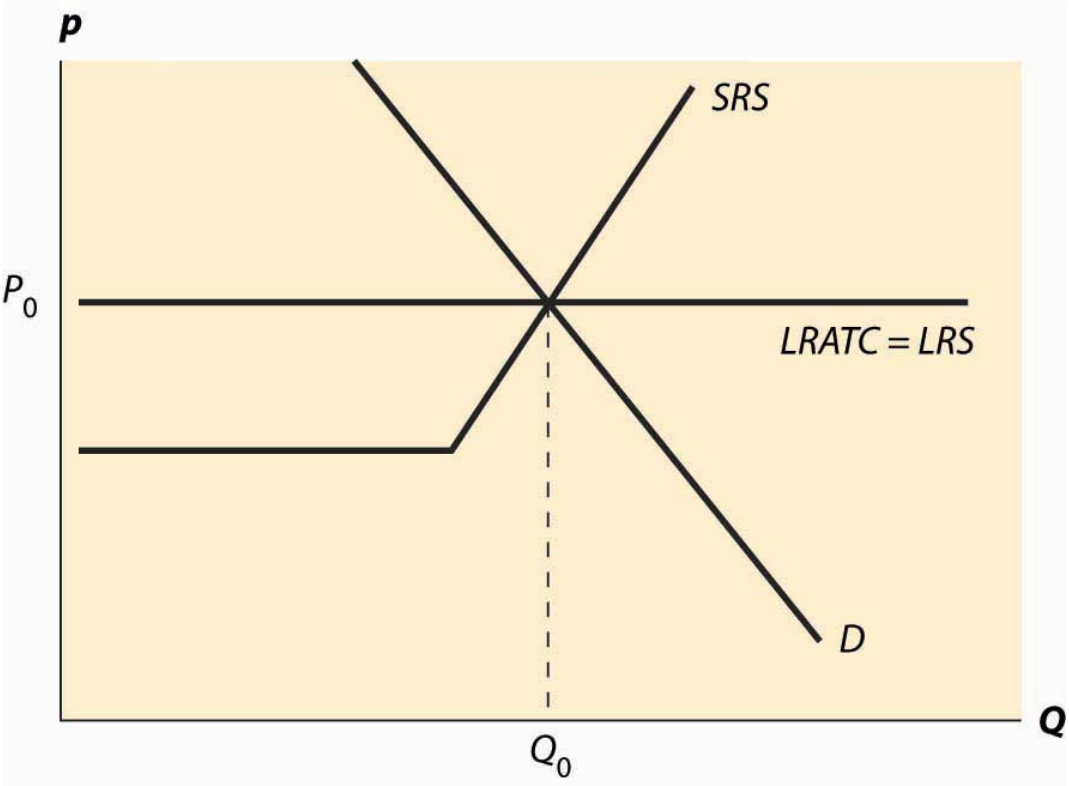


Figure 10.4, there is no long-run scale effect. The long-run average total cost is also the long-run industry supply.This may seem confusing, because supply is generally the *marginal* cost, not the average cost. However, because a firm will quit producing in the long-term if price falls below its minimum average cost, the long-term supply is just the minimum average cost of the individual firms because this is the marginal cost of the industry.

As drawn, the industry is in equilibrium, with price equal to P_0, which is the long-run average total cost, and also equates short-run supply and demand. That is, at the price of P_0, and industry output of Q_0, no firm wishes to shut down, no firm can make positive profits from entering, there is no excess output, and no consumer is rationed. Thus, no market participant has an incentive to change his or her behavior, so the market is in both long-run and short-run equilibrium. In **long-run equilibrium**, long-run demand equals long-run supply, and short-run demand equals short-run supply, so the market is also in **short-run equilibrium**, where short-run demand equals short-run supply.

Now consider an increase in demand. Demand might increase because of population growth, or because a new use for an existing product is developed, or because of income growth, or because the product becomes more useful. For example, the widespread adoption of the Atkins diet increased demand for high-protein products like beef jerky and eggs. Suppose that the change is expected to be permanent. This is important because the decision of a firm to enter is based more on expectations of future demand than on present demand.

Figure 10.5 reproduces the equilibrium figure, but with the curves “grayed out” to indicate a starting position and a darker, new demand curve, labeled D_1.

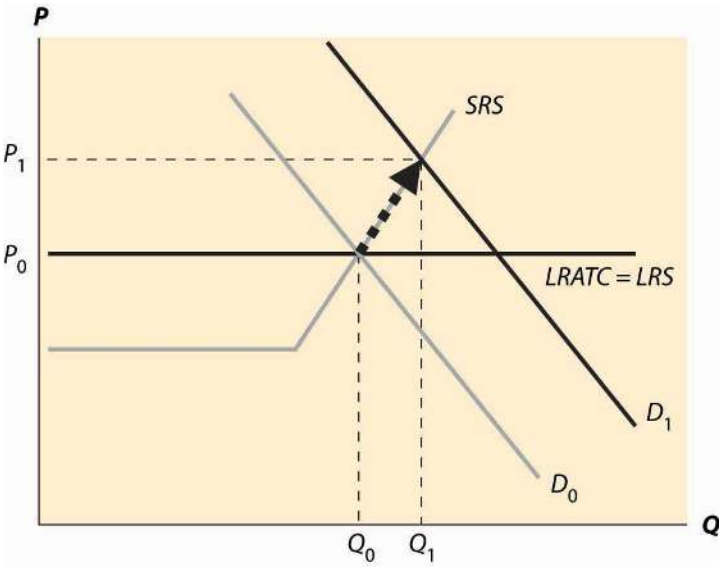


Figure 10.5 A shift in demand

The initial effect of the increased demand is that the price is bid up, because there is excess demand at the old price, P_0. This is reflected by a change in both price and quantity to P_1 and Q_1, to the intersection of the short-run supply (SRS) and the new demand curve. This is a short-run equilibrium, and persists temporarily because, in the short run, the cost of additional supply is higher.

At the new, short-run equilibrium, price exceeds the long-run supply (LRS) cost. This higher price attracts new investment in the industry. It takes some time for this new investment to increase the quantity supplied, but over time the new investment leads to increased output, and a fall in the price, as illustrated in Figure 10.6.

As new investment is attracted into the industry, the short-run supply shifts to the right because, with the new investment, more is produced at any given price level. This is illustrated with the darker short-run supply, SRS_2. The increase in price causes the price to fall back to its initial level and the quantity to increase still further to Q_2.

Figure 10.6 Return to long-run equilibrium

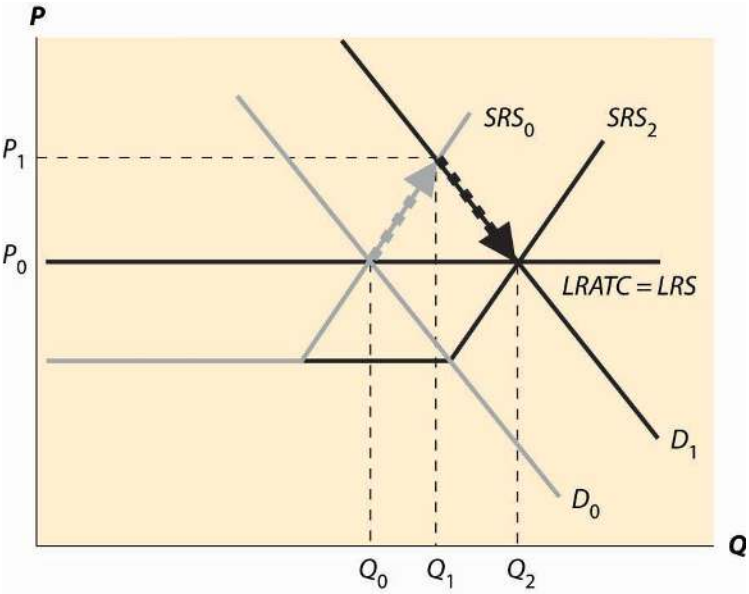


Figure 10.7.

In Figure 10.7, we start at the long-run equilibrium where LRS and D_0 and SRS_0 all intersect. If demand falls to D_1, the price falls to the intersection of the new demand and the old short-run supply, along SRS_0. At that point, exit of firms reduces the short-run supply and the price rises, following along the new demand D_1.

Figure 10.7 A decrease in demand

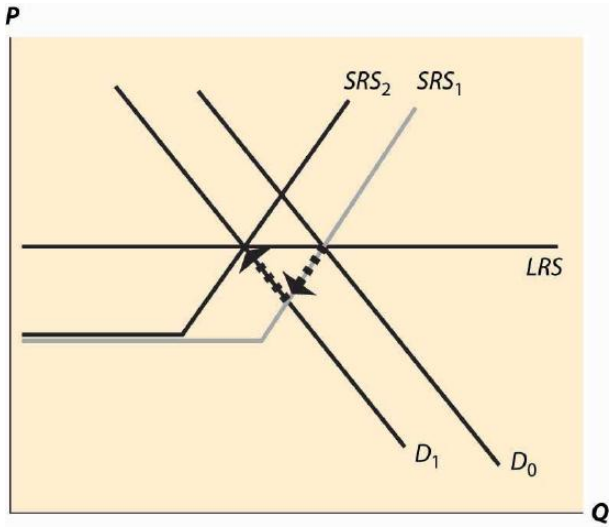


Figure 10.8, the fall in demand from D_0 to D_1 is sufficient to push the price to minimum average variable cost, which is the shutdown point of suppliers. Enough suppliers have to shut down to keep the price at this level, which induces a shift of the short-run supply, to SRS_1. Then there is additional shutdown, shifting in the short-run supply still further, but driving up the price (along the demand curve) until the long-term equilibrium is reached.

Figure 10.8 A big decrease in demand

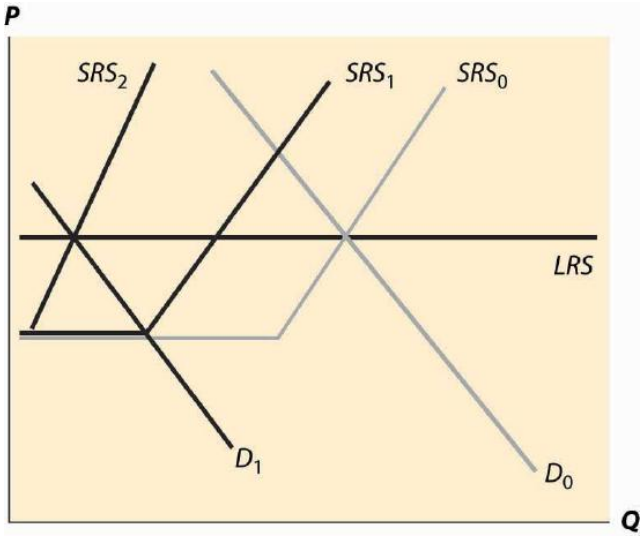


Figure 10.9. The increased costs reduce both the short-run supply (prices have to be higher in order to produce the same quantity) and the long-run supply. The short-run supply shifts upward to SRS_1 and the long-run supply to LRS_2. The short-run effect is to move to the intersection of the short-run supply and demand, which is at the price P_1 and the quantity Q_1. This price is below the long-run average cost, which is the long-run supply, so over time some firms don't replace their capital and there is disinvestment in the industry. This disinvestment causes the short-run supply to be reduced (move left) to SRS_2.

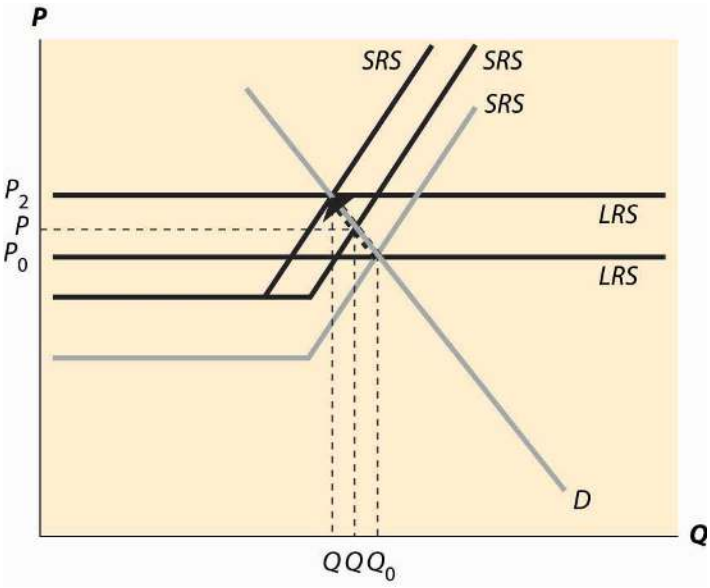


Figure 10.9 A decrease in supply

The case of a change in supply is more challenging because both the long-run supply and the short-run supply are shifted. But the logic—start at a long-run equilibrium, then look for the intersection of current demand and short-run supply, then look for the intersection of current demand and long-run supply—is the same whether demand or supply have shifted.

Key Takeaways

- A long-run equilibrium occurs at a price and quantity when the demand equals the long-run supply, and the number of firms is such that the short-run supply equals the demand.
- At long-run equilibrium prices, no firm wishes to shut down, no firm can make positive profits from entering, there is no excess output, and no consumer is rationed.
- An increase in demand to a system in long-run equilibrium first causes a short-run increase in output and a price increase. Then, because entry is profitable, firms enter. Entry shifts out short-run supply until the system achieves long-run equilibrium, decreasing prices back to their original level and increasing output.
- A decrease in demand creates a short-run equilibrium where existing short-run supply equals demand, with a fall in price and output. If the price fall is large enough (to average variable cost), some firms shut down. Then as firms exit, supply contracts, prices rise, and quantity contracts further.
- The case of a change in supply is more challenging because both the long-run supply and the short-run supply are shifted.

This page titled 10.3: Dynamics With Constant Costs is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

10.4: General Long-run Dynamics 10.4: General Long-run DynamicsEdit Update titlecancel

Learning Objectives

- If long-run costs aren't constant, how do changes in demand or costs affect short- and long-run prices and quantities traded?

The previous section made two simplifying assumptions that won't hold in all applications of the theory. First, it assumed constant returns to scale, so that long-run supply is horizontal. A perfectly elastic long-run supply means that price always eventually returns to the same point. Second, the theory didn't distinguish long-run from short-run demand. But with many products, consumers will adjust more over the long-term than immediately. As energy prices rise, consumers buy more energy-efficient cars and appliances, reducing demand. But this effect takes time to be seen, as we don't immediately scrap our cars in response to a change in the price of gasoline. The short-run effect is to drive less in response to an increase in the price, while the long-run effect is to choose the appropriate car for the price of gasoline.

To illustrate the general analysis, we start with a long-run equilibrium. Figure 10.10 reflects a long-run economy of scale, because the long-run supply slopes downward, so that larger volumes imply lower cost. The system is in long-run equilibrium because the short-run supply and demand intersection occurs at the same price and quantity as the long-run supply and demand intersection. Both short-run supply and short-run demand are less elastic than their long-run counterparts, reflecting greater substitution possibilities in the long run.

Figure 10.10 Equilibrium with external scale economy

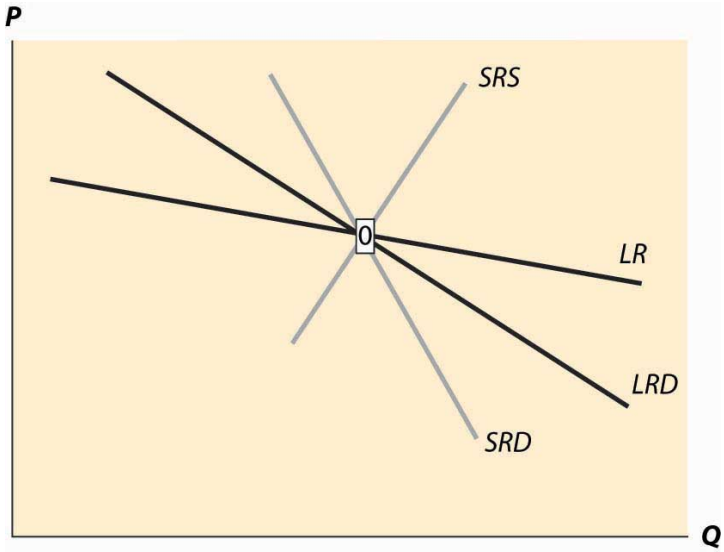


Figure 10.11 Decrease in demand

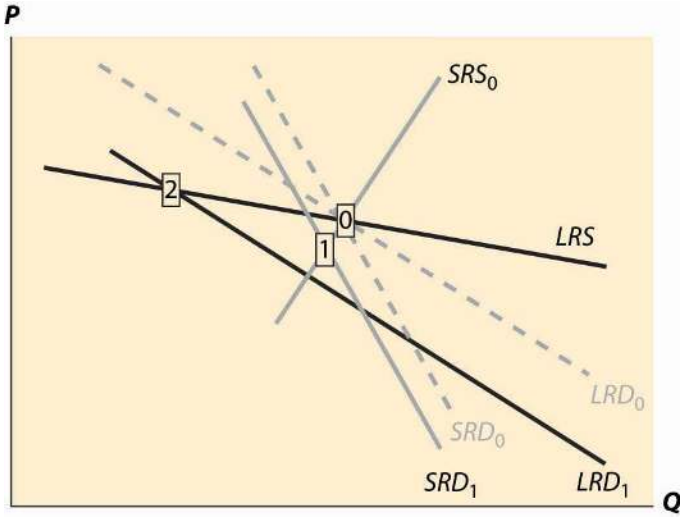


Figure 10.11. To reduce the proliferation of curves, we colored the old demand curves very faintly and marked the initial long-run equilibrium with a zero inside a small rectangle.The short-run demand and long-run demand have been shifted down by the same amount; that is, both reflect an equal reduction in value. This kind of shift might arise if, for instance, a substitute had become cheaper; but the equal reduction is not essential to the theory. In addition, the fact of equal reductions often isn't apparent from the diagram, because of the different slopes—to most observers, it appears that short-run demand fell less than long-run demand. This isn't correct, however, and one can see this because the intersection of the new short-run demand and long-run demand occurs directly below the intersection of the old curves, implying both fell by equal amounts. The intersection of short-run supply and short-run demand is marked with the number 1. Both long-run supply and long-run demand are more elastic than their short-run counterparts, which has an interesting effect. The short-run demand tends to shift down over time, because the price associated with the short-run equilibrium is above the long-run demand price for the short-run equilibrium quantity. However, the price associated with the short-run equilibrium is below the long-run supply price at that quantity. The effect is that buyers see the price as too high, and are reducing their demand, while sellers see the price as too low, and so are reducing their supply. Both short-run supply and short-run demand fall, until a long-run equilibrium is achieved.

Figure 10.12 Long-run after a decrease in demand

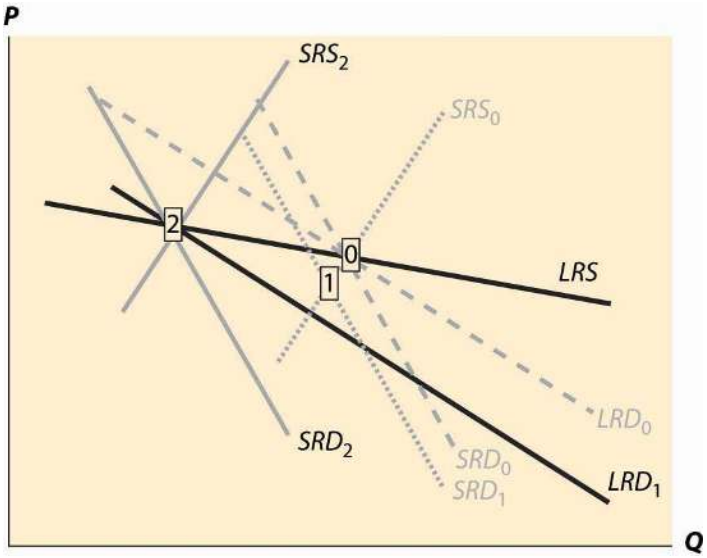


Figure 10.12 illustrates. The new long-run equilibrium has short-run demand and supply curves associated with it, and the system is in long-run equilibrium because the short-run demand and supply, which determine the current state of the system, intersect at the same point as the long-run demand and supply, which determine where the system is heading.

There are four basic permutations of the dynamic analysis—demand increase or decrease and a supply increase or decrease. Generally, it is possible for long-run supply to slope down—this is the case of an economy of scale—and for long-run demand to slope up.The demand situation analogous to an economy of scale in supply is a *network externality*, in which the addition of more users of a product increases the value of the product. Telephones are a clear example—suppose you were the only person with a phone—but other products like computer operating systems and almost anything involving adoption of a standard represent examples of network externalities. When the slope of long-run demand is greater than the slope of long-run supply, the system will

tend to be inefficient, because an increase in production produces higher average value and lower average cost. This usually means that there is another equilibrium at a greater level of production. This gives 16 variations of the basic analysis. In all 16 cases, the procedure is the same. Start with a long-run equilibrium and shift both the short-run and long-run levels of either demand or supply. The first stage is the intersection of the short-run curves. The system will then go to the intersection of the long-run curves.

An interesting example of competitive dynamics' concepts is the computer memory market, which was discussed previously. Most of the costs of manufacturing computer memory are fixed costs. The modern DRAM plant costs several billion dollars; the cost of other inputs—chemicals, energy, labor, silicon wafers—are modest in comparison. Consequently, the short-run supply is vertical until prices are very, very low; at any realistic price, it is optimal to run these plants 100% of the time.The plants are expensive, in part, because they are so clean—a single speck of dust falling on a chip ruins the chip. The Infineon DRAM plant in Virginia stopped operations only when a snowstorm prevented workers and materials from reaching the plant. The nature of the technology has allowed manufacturers to cut the costs of memory by about 30% per year over the past 40 years, demonstrating that there is a strong economy of scale in production. These two features—vertical short-run supply and strong economies of scale—are illustrated in Figure 10.13. The system is started at the point labeled with the number 0, with a relatively high price, and technology that has made costs lower than this price. Responding to the profitability of DRAM, short-run supply shifts out (new plants are built and die-shrinks permit increasing output from existing plants). The increased output causes prices to fall relatively dramatically because short-run demand is inelastic, and the system moves to the point labeled 1. The fall in profitability causes DRAM investment to slow, which allows demand to catch up, boosting prices to the point labeled 2. (One should probably think of Figure 10.13 as being in a logarithmic scale.)

Figure 10.13 DRAM market

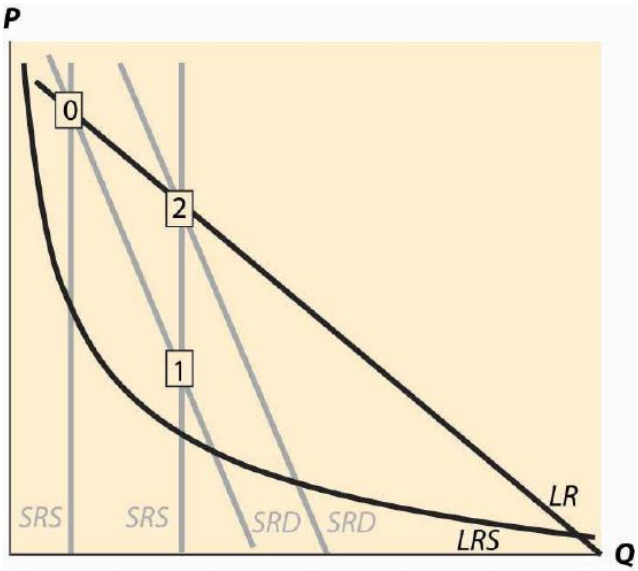


Figure 10.14, which shows DRAM industry revenues in billions of dollars from 1992 to 2003, and projections of 2004 and 2005.Two distinct data sources were used, which is why there are two entries for each of 1998 and 1999.

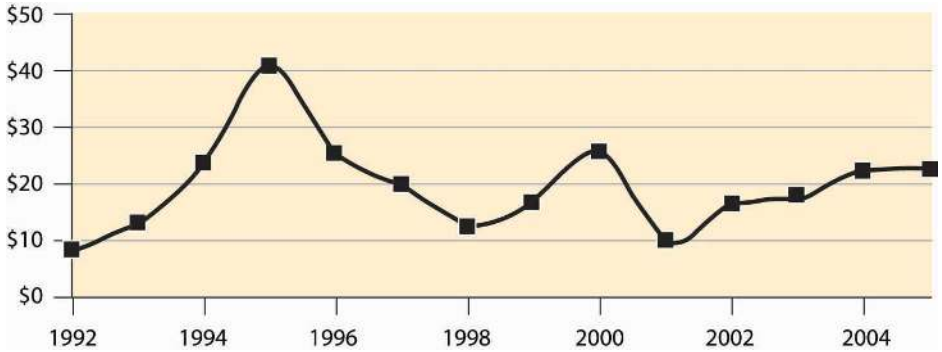


Figure 10.14 DRAM revenue cycle

KEY TAKEAWAY

- In general, both demand and supply may have long-run and short-run curves. In this case, when something changes, initially the system moves to the intersections of the current short-run supply and demand for a short-run equilibrium, then to the intersection of the long-run supply and demand. The second change involves shifting short-run supply and demand curves.

EXERCISES

1. Land close to the center of a city is in fixed supply, but it can be used more intensively by using taller buildings. When the population of a city increases, illustrate the long- and short-run effects on the housing markets using a graph.
2. Emus can be raised on a wide variety of ranch land, so that there are constant returns to scale in the production of emus in the long run. In the short run, however, the population of emus is limited by the number of breeding pairs of emus, and the supply is essentially vertical. Illustrate the long- and short-run effects of an increase in demand for emus. (In the late 1980s, there was a speculative bubble in emus, with prices reaching $80,000 per breeding pair, in contrast to $2,000 or so today.)
3. There are long-run economies of scale in the manufacture of computers and their components. There was a shift in demand away from desktop computers and toward notebook computers around the year 2001. What are the short- and long-run effects? Illustrate your answer with two diagrams, one for the notebook market and one for the desktop market. Account for the fact that the two products are substitutes, so that if the price of notebook computers rises, some consumers shift to desktops. (To answer this question, start with a time 0 and a market in long-run equilibrium. Shift demand for notebooks out and demand for desktops in.) What happens in the short run? What happens in the long run to the prices of each? What does that price effect do to demand?

This page titled 10.4: General Long-run Dynamics is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

11: Investment

11.1: Present Value
11.2: Investment
11.3: Investment Under Uncertainty
11.4: Resource Extraction
11.5: A Time to Harvest
11.6: Collectibles
11.7: Summer Wheat

This page titled 11: Investment is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

11.1: Present Value 11.1: Present ValueEdit Update titlecancel

Learning Objectives

- How is a stream of payments or liabilities evaluated?

The promise of $1 in the future is not worth $1 today. There are a variety of reasons why a promise of future payments is not worth the face value today, some of which involve risk that the money may not be paid. Let's set aside such risk for the moment. Even when the future payment is perceived to occur with negligible risk, most people prefer $1 today to $1 payable a year hence. One way to express this is by the **present value**: The value today of a future payment of a dollar is less than a dollar. From a present value perspective, future payments are **discounted**.

From an individual perspective, one reason that one should value a future payment less than a current payment is due to **arbitrage**.Arbitrage is the process of buying and selling in such a way as to make a profit. For example, if wheat is selling for $3 per bushel in New York but $2.50 per bushel in Chicago, one can buy in Chicago and sell in New York, profiting by $0.50 per bushel minus any transaction and transportation costs. Such arbitrage tends to force prices to differ by no more than transaction costs. When these transaction costs are small, as with gold, prices will be about the same worldwide. Suppose you are going to need $10,000 one year from now to put a down payment on a house. One way of producing $10,000 is to buy a government bond that pays $10,000 a year from now. What will that bond cost you? At current interest rates, a secure bondEconomists tend to consider U.S. federal government securities secure, because the probability of such a default is very, very low. will cost around $9,700. This means that no one should be willing to pay $10,000 for a future payment of $10,000 because, instead, one can have the future $10,000 by buying the bond, and will have $300 left over to spend on cappuccinos or economics textbooks. In other words, if you will pay $10,000 for a secure promise to repay the $10,000 a year hence, then I can make a successful business by selling you the secure promise for $10,000 and pocketing $300.

This arbitrage consideration also suggests how to value future payments: discount them by the relevant interest rate.

Example (Auto loan): You are buying a $20,000 car, and you are offered the choice to pay it all today in cash, or to pay $21,000 in one year. Should you pay cash (assuming you have that much in cash) or take the loan? The loan is at a 5% annual interest rate because the repayment is 5% higher than the loan amount. This is a good deal for you if your alternative is to borrow money at a higher interest rate; for example, on (most) credit cards. It is also a good deal if you have savings that pay more than 5%—if buying the car with cash entails cashing in a certificate of deposit that pays more than 5%, then you would be losing the difference. If, on the other hand, you are currently saving money that pays less than 5% interest, paying off the car is a better deal.

The formula for present value is to discount by the amount of interest. Let's denote the interest rate for the next year as r_1, the second year's rate as r_2, and so on. In this notation, $1 invested would pay $\$(1+r_1)$ next year, or $\$(1+r_1)\times(1+r_2)$ after 2 years, or $\$(1+r_1)\times(1+r_2)\times(1+r_3)$ after 3 years. That is, r_i is the interest rate that determines the value, at the end of year *i*, of $1 invested at the start of year *i*. Then if we obtain a stream of payments A_0 immediately, A_1 at the end of year one, A_2 at the end of year two, and so on, the present value of that stream is

$$\mathrm{PV} = \mathrm{A0} + \mathrm{A11} + \mathrm{r1} + \mathrm{A2}(1+\mathrm{r1})(1+\mathrm{r2}) + \mathrm{A2}(1+\mathrm{r1})(1+\mathrm{r2})(1+\mathrm{r3}) + \ldots$$

Example (**Consolidated annuities or consols**): What is the value of $1 paid at the end of each year forever, with a fixed interest rate *r*? Suppose the value is *v*. ThenThis development uses the formula that, for $-1 < a < 1, 11 - a = 1 + a + a2 + \ldots$, which is readily verified. Note that this formula involves an infinite series.

$$v = 11 + r + 1(1+r)2 + 1(1+r)3 + \ldots = 11 - 11 + r - 1 = 1r.$$

At a 5% interest rate, $1 million per year paid forever is worth $20 million today. Bonds that pay a fixed amount every year forever are known as consols; no current government issues consols.

Example (Mortgages): Again, fix an interest rate *r*, but this time let *r* be the monthly interest rate. A mortgage implies a fixed payment per month for a large number of months (e.g., 360 for a 30-year mortgage). What is the present value of these payments over *n* months? A simple way to compute this is to use the consol value, because

$$\begin{aligned} M &= 11 + r + 1(1+r)2 + 1(1+r)3 + \ldots + 1(1+r)n = 1r - 1(1+r)n + 1 - 1(1+r)n + 2 - 1(1+r)n + 3 - \ldots \\ &= 1r - 1(1+r)n(1(1+r) + 1(1+r)2 + 1(1+r)3 + \ldots) = 1r - 1(1+r)n1r = 1r(1 - 1(1+r)n). \end{aligned}$$

Thus, at a monthly interest rate of 0.5%, paying $1 per month for 360 months produces a present value *M* of 1 0.005 (1− 1 (1.005) 360)=$166.79 . Therefore, to borrow $100,000, one would have to pay $100,000 $166.79 =$599.55 per month. It is important to remember that a different loan amount just changes the scale: Borrowing $150,000 requires a payment of $150,000 $166.79 =$899.33 per month, because $1 per month generates $166.79 in present value.

Example (Simple and compound interest): In the days before calculators, it was a challenge to actually solve interest-rate formulas, so certain simplifications were made. One of these was **simple interest**, which meant that daily or monthly rates were translated into annual rates by incorrect formulas. For example, with an annual rate of 5%, the simple interest daily rate is 5% 365 =0.07692%. The fact that this is incorrect can be seen from the calculation (1+ .05 365) 365 =1.051267%, which is the **compound interest** calculation. Simple interest increases the annual rate, so it benefits lenders and harms borrowers. (Consequently, banks advertise the accurate annual rate on savings accounts—when consumers like the number to be larger—but not on mortgages, although banks are required by law to disclose, but not to advertise widely, actual annual interest rates on mortgages.)

Example (Obligatory lottery): You win the lottery, and the paper reports that you've won $20 million. You're going to be paid $20 million, but is it worth $20 million? In fact, you get $1 million per year for 20 years. However, in contrast to our formula, you get the first million right off the bat, so the value is

$$\mathrm{PV} = 1 + 11 + \mathrm{r} + 1(1 + \mathrm{r})2 + 1(1 + \mathrm{r})3 + \ldots + 1(1 + \mathrm{r})19 = 1 + 1\mathrm{r}(1 - 1(1 + \mathrm{r})19)$$

Table 11.1 computes the present value of our $20 million dollar lottery, listing the results in thousands of dollars, at various interest rates. At 10% interest, the value of the lottery is less than half the "number of dollars" paid; and even at 5%, the value of the stream of payments is 65% of the face value.

Table 11.1 Present value of $20 million

r	3%	4%	5%	6%	7%	10%
PV (000s)	$15,324	$14,134	$13,085	$12,158	$11,336	$9,365

The lottery example shows that interest rates have a dramatic impact on the value of payments made in the distant future. Present value analysis is the number one tool used in MBA programs, where it is known as **net present value, or NPV**, analysis. It is accurate to say that the majority of corporate investment decisions are guided by an NPV analysis.

Example (Bond prices): A standard **Treasury bill** has a fixed future value. For example, it may pay $10,000 in one year. It is sold at a discount off the face value, so that a one-year, $10,000 bond might sell for $9,615.39, producing a 4% interest rate. To compute the effective interest rate *r*, the formula relating the future value *FV*, the number of years *n*, and the price is

$$(1 + r)n = FVPrice$$

or

$$r = (FVPrice)1n - 1.$$

We can see from either formula that Treasury bill prices move inversely to interest rates—an increase in interest rates reduces Treasury prices. Bonds are a bit more complicated. Bonds pay a fixed interest rate set at the time of issue during the life of the bond, generally collected semiannually, and the face value is paid at the end of the term. These bonds were often sold on long terms, as much as 30 years. Thus, a three-year, $10,000 bond at 5% with semiannual payments would pay $250 at the end of each half year for 3 years, and pay $10,000 at the end of the 3 years. The net present value, with an annual interest rate *r*, is

$$\mathrm{NPV} = \$250(1 + r)12 + \$250(1 + r)1 + \$250(1 + r)32 + \$250(1 + r)2 + \$250(1 + r)52 + \$250(1 + r)3 + \$10000(1 + r)3$$

The net present value will be the price of the bond. Initially, the price of the bond should be the face value, since the interest rate is set as a market rate. The U.S. Treasury quit issuing such bonds in 2001, replacing them with bonds in which the face value is paid and then interest is paid semiannually.

Key Takeaways

- Capital goods change slowly, in part because they are durable.
- The acquisition of goods that will be used over time, whether they be factories, homes, or televisions, is known as investment.
- The promise of $1 in the future is not worth $1 today. The difference is a discount on future payments.
- Arbitrage involves buying and selling such that a positive surplus is earned.

- Arbitrage is possible unless future payments are discounted by the appropriate interest rate.
- “Simple” interest means that daily, monthly, or annual rates are translated into daily, monthly, or annual rates by incorrect formulas. Accurate calculations are known as compound interest.
- A standard Treasury bill has a fixed future value. Treasury bill prices move inversely to interest rates—an increase in interest rates reduces Treasury prices.
- Bonds pay a fixed interest rate set at the time of issue during the life of the bond, generally collected semiannually, and the face value is paid at the end of the term.

EXERCISES

1. At a 7% annual interest rate, what is the present value of $100 paid at the end of one year, and $200 paid at the end of the second year?
2. Compute the NPV of the 3-year, $10,000 bond with $250 semiannual payments semiannually, at an interest rate of 4%.
3. You can finance your $20,000 car with a straight 5% loan paid monthly over 5 years, or a loan with one year interest-free followed by 4 years of 7% interest. (Hint: In both cases, figure out the fixed monthly payments that produce an NPV equal to $20,000.)
4. You win the lottery. At what interest rate should you accept $7 million today over 20 annual payments of $500,000?
5. An investor discounts future profits at 5% per year. Suppose a stock pays $1 in dividends after one year, growing 1% each year thereafter. How much should the stock be worth today?
6. You are buying a $20,000 car. If you make monthly payments of $1,000, how long will it take you to pay off the debt if the interest rate is 1% per month? How does this change when the interest rate drops to 0.5%?

This page titled 11.1: Present Value is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

11.2: Investment 11.2: InvestmentEdit Update titlecancel

Learning Objectives

- How do I evaluate an investment opportunity?

A simple investment project requires an investment, *I*, followed by a return over time. If you dig a mine, drill an oil well, build an apartment building or a factory, or buy a share of stock, you spend money now, hoping to earn a return in the future. We will set aside the very important issue of risk until the next subsection, and ask how one makes the decision to invest.

The NPV approach involves assigning a rate of return *r* that is reasonable for the specific project and then computing the corresponding present value of the expected stream of payments. Since the investment is initially expended, it is counted as negative revenue. This yields an expression that looks like

$$\mathrm{NPV} = -\mathrm{I} + \mathrm{R11} + \mathrm{r} + \mathrm{R2}(1+\mathrm{r})2 + \mathrm{R3}(1+\mathrm{r})3 + ...$$

where R_1 represents first-year revenues, R_2 represents second-year revenues, and so on.The most common approach treats revenues within a year as if they are received at the midpoint, and then discounts appropriately for that mid-year point. The present discussion abstracts from this practice. The investment is then made when NPV is positive—because this would add to the net value of the firm.

Carrying out an NPV analysis requires two things. First, investment and revenues must be estimated. This is challenging, especially for new products where there is no direct way of estimating demand, or with uncertain outcomes like oil wells or technological research.The building of the famed Sydney Opera House, which looks like billowing sails over Sydney Harbor in Australia, was estimated to cost $7 million and actually cost $105 million. A portion of the cost overrun was due to the fact that the original design neglected to install air conditioning. When this oversight was discovered, it was too late to install a standard unit, which would interfere with the excellent acoustics, so instead an ice hockey floor was installed as a means of cooling the building. Second, an appropriate rate of return must be identified. The rate of return is difficult to estimate, mostly because of the risk associated with the investment payoffs. Another difficulty is recognizing that project managers have an incentive to inflate the payoffs and minimize the costs to make the project appear more attractive to upper management. In addition, most corporate investment is financed through retained earnings, so that a company that undertakes one investment is unable to make other investments, so the interest rate used to evaluate the investment should account for opportunity cost of corporate funds. As a result of these factors, interest rates of 15%–20% are common for evaluating the NPV of projects of major corporations.

Example (Silver mine): A company is considering developing a silver mine in Mexico. The company estimates that developing the mine requires building roads and opening a large hole in the ground, which would cost $4 million per year for 4 years, during which time the mines generates zero revenue. Starting in year 5, the expenses would fall to $2 million per year, and $6 million in net revenue would accrue from the silver that is mined for the next 40 years. If the company cost of funds is 18%, should it develop the mine?

The earnings from the mine are calculated in the table below. First, the NPV of the investment phase during years 0, 1, 2, and 3 is

$$\mathrm{NPV} = -4 + -41.18 + -4(1.18)2 + -4(1.18)3 = -12.697$$

A dollar earned in each of years 4 through 43 has a present value of

$$1(1+r)4 + 1(1+r)5 + 1(1+r)6 + ... + 1(1+r)43 = 1(1+r)3 \times 1r(1-1(1+r)40) = 13.377$$

The mine is just profitable at 18%, in spite of the fact that its $4 million payments are made in 4 years, after which point $4 million in revenue is earned for 40 years. The problem in the economics of mining is that 18% makes the future revenue have quite modest present values.

Year	Earnings ($M)/yr	*PV* ($M)
0–3	–4	–12.697
4–43	4	13.377
Net		0.810

There are other approaches for deciding to take an investment. In particular, the **internal rate of return (IRR)** approach solves the equation $NPV = 0$ for the interest rate. Then the project is undertaken if the rate of return is sufficiently high. This approach is flawed because the equation may have more than one solution—or no solutions—and the right thing to do in these events is not transparent. Indeed, the IRR approach gets the profit-maximizing answer only if it agrees with NPV. A second approach is the payback period, which asks calculates the number of years a project must be run before profitability is reached. The problem with the payback period is deciding between projects—if I can only choose one of two projects, the one with the higher NPV makes the most money for the company. The one with the faster payback may make a quite small amount of money very quickly, but it isn't apparent that this is a good choice. When a company is in risk of bankruptcy, a short payback period might be valuable, although this would ordinarily be handled by employing a higher interest rate in an NPV analysis. NPV does a good job when the question is whether or not to undertake a project, and it does better than other approaches to investment decisions. For this reason, NPV has become the most common approach to investment decisions. Indeed, NPV analysis is more common than all other approaches combined. NPV does a poor job, however, when the question is whether to undertake a project or to delay the project. That is, NPV answers "yes or no" to investment, but when the choice is "yes or wait," NPV requires an amendment.

Key Takeaways

- The NPV approach involves assigning a rate of return r that is reasonable for, and specific to, the project and then computing the present value of the expected stream of payments. The investment is then made when NPV is positive—since this would add to the net value of the firm.
- Carrying out an NPV analysis requires estimating investment and revenues and identifying an appropriate rate of return.
- Interest rates of 15%–20% are common for evaluating the NPV of projects of major corporations.

EXERCISES

1. Suppose that, without a university education, you'll earn $25,000 per year. A university education costs $20,000 per year, and you forgo the $25,000/year that you would have earned for 4 years. However, you earn $50,000 per year for the following 40 years. At 7%, what is the NPV of the university education?
2. Now that you've decided to go to the university based on the previous answer, suppose that you can attend East State U, paying $3,000 per year for 4 years and earning $40,000 per year when you graduate, or you can attend North Private U, paying $22,000 per year for the 4 years and earning $50,000 per year when you graduate. Which is the better deal at 7%?
3. A bond is a financial instrument that pays a fixed amount, called the face value, at a maturity date. Bonds can also pay out fixed payments, called coupons, in regular intervals up until the maturity date. Suppose a bond with face value $1,000 sells for $900 on the market and has annual coupon payments starting a year from today up until its maturity date 10 years from now. What is the coupon rate? Assume $r = 10\%$.
4. The real return on stocks averages about 4% annually. Over 40 years, how much will $1,000 invested today grow?
5. You have made an invention. You can sell the invention now for $1 million and work at something else, producing $75,000 per year for 10 years. (Treat this income as received at the start of the year.) Alternatively, you can develop your invention, which requires working for 10 years, and it will net you $5 million in 10 years hence. For what interest rates are you better off selling now? (Please approximate the solution.)
6. A company is evaluating a project with a start-up fee of $50,000 but pays $2,000 every second year thereafter, starting 2 years from now. Suppose that the company is indifferent about taking on the project—or not. What discount rate is the company using?

This page titled 11.2: Investment is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

11.3: Investment Under Uncertainty 11.3: Investment Under UncertaintyEdit Update titlecancel

Learning Objectives

- What is the effect of risk on investment?
- What is an option?

Risk has a cost, and people and corporations buy insurance against financial risk.For example, NBC spent $6 million to buy an insurance policy against U.S. nonparticipation in the 1980 Moscow Summer Olympic Games—and the United States didn't participate (because of the Soviet invasion of Afghanistan)—and NBC was paid $94 million from the policy. The standard approach to investment under uncertainty is to compute an NPV, using a "risk-adjusted" interest rate to discount the expected values received over time. The interest rate is increased or lowered depending on how risky the project is.

For example, consider a project like oil exploration. The risks are enormous. Half of all underwater tracts in the Gulf Coast near Louisiana and Texas that are leased are never drilled, because they turn out to be a bad bet. Half of all the tracts that are drilled are dry. Hence, three quarters of the tracts that are sold produce zero or negative revenue. To see how the economics of such a risky investment might be developed, suppose that the relevant rate of return for such investments is 18%. Suppose further that the tract can be leased for $500,000 and the initial exploration costs $1 million. If the tract has oil (with a 25% probability), it produces $1 million per year for 20 years and then runs dry. This gives an expected revenue of $250,000 per year. To compute the expected net present value, we first compute the returns:

Table 11.2 Oil tract return

	Expected revenue	EPV
0	–$1.5M	–$1.5M
1–20	$0.25M	$1.338M
Net		–$0.162

At 18%, the investment is a loss—the risk is too great given the average returns.

A very important consideration for investment under uncertainty is the choice of interest rate. It is crucial to understand that the interest rate is specific to the project and not to the investor. This is perhaps the most important insight of corporate finance generally: The interest rate should adjust for the risk associated with the project and not for the investor. For example, suppose hamburger retailer McDonald's is considering investing in a cattle ranch in Peru. McDonald's is overall a very low-risk firm, but this particular project is quite risky because of local conditions. McDonald's still needs to adjust for the market value of the risk it is undertaking, and that value is a function of the project risk, not the risk of McDonald's other investments.

This basic insight of **corporate finance** (the study of funding of operations of companies)—the appropriate interest rate is determined by the project, not by the investor—is counterintuitive to most of us because it doesn't apply to our personal circumstances. For individuals, the cost of borrowing money is mostly a function of our own personal circumstances, and thus the decision of whether to pay cash for a car or borrow the money is not so much a function of the car that is being purchased but of the wealth of the borrower. Even so, personal investors borrow money at distinct interest rates. Mortgage rates on houses are lower than interest rates on automobiles, and interest rates on automobiles are lower than on credit cards. This is because the "project" of buying a house has less risk associated with it: The percentage loss to the lender in the event of borrower default is lower on a house than on a car. Credit cards carry the highest interest rates because they are unsecured by any asset.

One way of understanding why the interest rate is project-specific, but not investor-specific, is to think about undertaking the project by creating a separate firm to make the investment. The creation of subsidiary units is a common strategy, in fact. This subsidiary firm created to operate a project has a value equal to the NPV of the project using the interest rate specific to the subsidiary, which is the interest rate for the project, independent of the parent. For the parent company, owning such a firm is a good thing if the firm has positive value, but not otherwise.It may seem that synergies between parent and subsidiary are being neglected here, but synergies should be accounted for at the time they produce value—that is, as part of the stream of revenues of the subsidiary.

Investments in oil are subject to another kind of uncertainty: price risk. The price of oil fluctuates. Moreover, oil pumped and sold today is not available for the future. Should you develop and pump the oil you have today, or should you hold out and sell in the future? This question, known as the option value of investment, is generally somewhat challenging and arcane, but a simple example provides a useful insight. An **option** is the right to buy or sell at a price determined in advance.

To develop this example, let's set aside some extraneous issues first. Consider a very simple investment, in which either C is invested or not.This theory is developed in striking generality by Avinash Dixit and Robert Pindyck, *Investment Under Uncertainty*, Princeton University Press, 1994. If C is invested, a value V is generated. The cost C is a constant; it could correspond to drilling or exploration costs or, in the case of a stock option, the **strike price** of the option, which is the amount one pays to obtain the share of stock. The value V, in contrast, varies from time to time in a random fashion. To simplify the analysis, we assume that V is uniformly distributed on the interval [0, 1], so that the probability of V falling in an interval

$$[a, b] is (b–a) \text{ if } 0 \le a \le b \le 1$$

. The option only has value if $C < 1$, which we assume for the rest of this section.

The first thing to note is that the optimal rule to make the investment is a cutoff value—that is, to set a level V_0 and exercise the option if, and only if, $V \ge V_0$. This is because—if you are willing to exercise the option and generate value V—you should be willing to exercise the option and obtain even more value. The NPV rule simply says $V_0 = C$; that is, invest whenever it is profitable. The purpose of the example developed below is to provide some insight into how far wrong the NPV rule will be when option values are potentially significant.

Now consider the value of option to invest, given that the investment rule $V \ge V_0$ is followed. Call this option value $J(V_0)$. If the realized value V exceeds V_0, one obtains $V – C$. Otherwise, one delays the investment, producing a discounted level of the same value. This logic says,

$$J(\vee 0) = (1 - \vee 0)(1 + \vee 02 - C) + V0(11 + r](\vee 0))$$

This expression for $J\,(V_0)$ is explained as follows. First, the hypothesized distribution of V is uniform on [0, 1]. Consequently, the value of V will exceed V_0 with probability $1 – V_0$. In this event, the expected value of V is the midpoint of the interval $[V_0, 1]$, which is $½(V_0 + 1)$. The value $½(V_0 + 1) – C$ is the average payoff from the strategy of investing whenever $V \ge V_0$, which is obtained with probability $1 – V_0$. Second, with probability V_0, the value falls below the cutoff level V_0. In this case, no investment is made and, instead, we wait until the next period. The expected profits of the next period are $J\,(V_0)$, and these profits are discounted in the standard way.

The expression for J is straightforward to solve:

$$\mathrm{J(V0) = (1 - V0)(1 + V02 - C)1 - V01 + r}$$

Rudimentary calculus shows

$$\mathrm{J'(V0) = 1 + 2rC + V02 - 2(1 + r) \vee 02(1 + r)(1 - V01 + r)2}$$

First, note that $\mathrm{J'(C) > 0 \text{ and } J'(1) < 0}$, which together imply the existence of a maximum at a value V_0 between C and 1, satisfying J ′ (V 0)=0. Second, the solution occurs at

$$V0 = (1 + r) - (1 + r)2 - (1 + 2rC) = (1 + r) - r2 + 2r(1 - C)$$

The positive root of the quadratic has $V_0 > 1$, which entails never investing, and hence is not a maximum. The profit-maximizing investment strategy is to invest whenever the value exceeds V_0 given by the negative root in the formula. There are a couple of notable features about this solution. First, at $r = 0$, $V_0 = 1$. This is because $r = 0$ corresponds to no discounting, so there is no loss in holding out for the highest possible value. Second, as $r \to \infty$, $V_0 \to C$. As $r \to \infty$, the future is valueless, so it is worth investing if the return is anything over costs. These are not surprising findings, but rather quite the opposite: They should hold in any reasonable formulation of such an investment strategy. Moreover, they show that the NPV rule, which requires $V_0 = C$, is correct only if the future is valueless.

How does this solution behave? The solution is plotted as a function of r, for $C = 0$, 0.25, and 0.5, in Figure 11.1.

The horizontal axis represents interest rates, so this figure shows *very* high interest rates by current standards, up to 200%. Even so, V_0 remains substantially above C. That is, even when the future has very little value, because two-thirds of the value is destroyed

by discounting each period, the optimal strategy deviates significantly from the NPV strategy. Figure 11.2 shows a close-up of this graph for a more reasonable range of interest rates, for interest rates of 0%–10%.

Figure 11.1 Investment strike price given interest rate *r* in percent

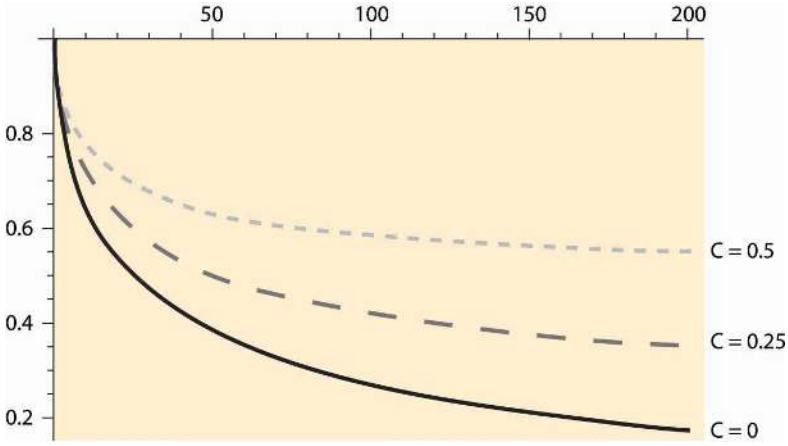


Figure 11.2 shows the cutoff values of investment for three values of *C*, the cost of the investment. These three values are 0 (lowest curve), 0.25 (the middle, dashed curve), and 0.5 (the highest, dotted line). Consider the lowest curve, with $C = 0$. The NPV of this project is always positive—there are no costs, and revenues are positive. Nevertheless, because the investment can only be made once, it pays to hold out for a higher level of payoff; indeed, for 65% or more of the maximum payoff. The economics at an interest rate of 10% is as follows. By waiting, there is a 65% chance that 10% of the potential value of the investment is lost. However, there is a 35% chance of an even higher value. The optimum value of V_0 trades these considerations off against each other.

For $C = 0.25$, at 10% the cutoff value for taking an investment is 0.7, nearly three times the actual cost of the investment. Indeed, the cutoff value incorporates two separate costs: the actual expenditure on the investment *C* and the lost opportunity to invest in the future. The latter cost is much larger than the expenditure on the investment in many circumstances and, in this example, can be quantitatively much larger than the actual expenditure on the investment.

Some investments can be replicated. There are over 13,000 McDonald's restaurants in the United States, and building another one doesn't foreclose building even more. For such investments, NPV analysis gets the right answer, provided that appropriate interest rates and expectations are used. Other investments are difficult to replicate or logically impossible to replicate—having pumped and sold the oil from a tract, that tract is now dry. For such investments, NPV is consistently wrong because it neglects the value of the option to delay the investment. A correct analysis adds a lost value for the option to delay the cost of the investment—a value that can be quantitatively large—as we have seen.

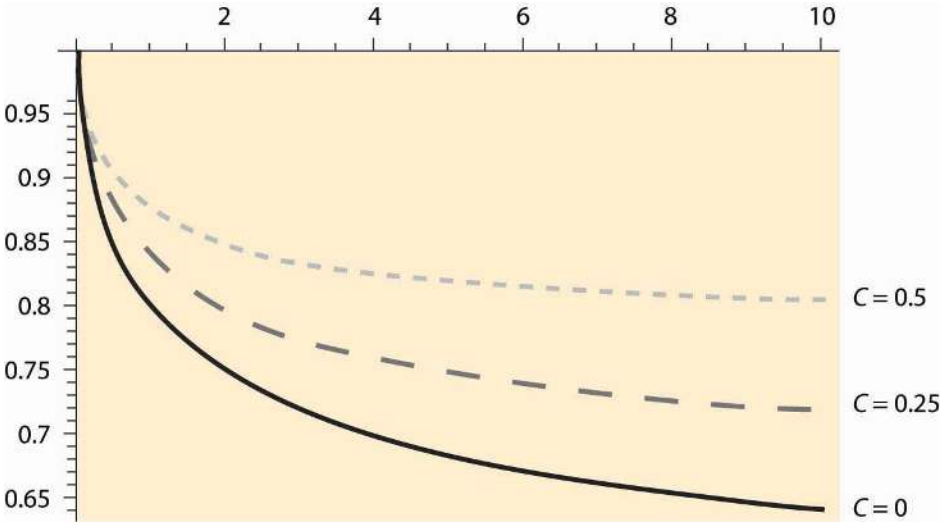


Figure 11.2 Investment strike price given interest rate *r* in percent

Example: When should you refinance a mortgage? Suppose you are paying 10% interest on a $100,000 mortgage, and it costs $5,000 to refinance; but refinancing permits you to lock in a lower interest rate, and hence pay less. When is it a good idea? To

answer this question, we assume that the $5,000 cost of refinancing is built into the loan so that, in essence, you borrow $105,000 at a lower interest rate when you refinance. This is actually the most common method of refinancing a mortgage.

To simplify the calculations, we will consider a mortgage that is never paid off; that is, one pays the same amount per year forever. If the mortgage isn't refinanced, one pays 10% of the $100,000 face value of the mortgage each year, or $10,000 per year. If one refinances at interest rate *r*, one pays *r* × $105,000 per year, so the NPV of refinancing is

$$\text{NPY} = \$10,000 - r \times \$105,000$$

Thus, NPV is positive whenever $r < 10105 = 9.52\%$

Should you refinance when the interest rate drops to this level? No. At this level, you would exactly break even, but you would also be carrying a $105,000 mortgage rather than a $100,000 mortgage, making it harder to benefit from any further interest-rate decreases. The only circumstance in which refinancing at 9.52% is sensible is if interest rates can't possibly fall further.

When should you refinance? That depends on the nature and magnitude of the randomness governing interest rates, preferences over money today versus money in the future, and attitudes to risk. The model developed in this section is not a good guide to answering this question, primarily because the interest rates are strongly correlated over time. However, an approximate guide to implementing the option theory of investment is to seek an NPV of twice the investment, which would translate into a refinance point of around 8.5%.

Key Takeaways

- The standard approach to investment under uncertainty is to compute an NPV, with the interest rate used adjusted to compensate for the risk.
- The most important thing to understand is that the interest rate is specific to the project and not to the investor.
- The option value of investment includes the value of decisions that have not yet been made. Building a factory today forecloses the opportunity of building the factory a year later, when better information concerning market conditions has been realized, but also creates the option of selling over the next year.
- NPV gets the right answer when investments can be replicated.
- An important example of the option value is refinancing a home.

EXERCISE

1. You are searching for a job. The net value of jobs that arise is uniformly distributed on the interval [0, 1]. When you accept a job, you must stop looking at subsequent jobs. If you can interview with one employer per week, what jobs should you accept? Use a 7% annual interest rate.

   (Hint: Relate the job-search problem to the investment problem, where accepting a job is equivalent to making the investment. What is *c* in the job-search problem? What is the appropriate interest rate?)

This page titled 11.3: Investment Under Uncertainty is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

11.4: Resource Extraction 11.4: Resource ExtractionEdit Update titlecancel

Learning Objectives

- How much of a limited resource should be consumed today, and how much should be saved for future consumption?

For the past 60 years, the world has been "running out of oil." There are news stories about the end of the reserves being only 10, 15, or 20 years away. The tone of these stories is that, at this time, we will run out of oil completely and prices will be extraordinarily high. Industry studies counter that more oil continues to be found and that the world is in no danger of running out of oil.

If you believe that the world will run out of oil, what should you do? You should buy and hold. That is, if the price of oil in 20 years is going to be $1,000 per barrel, then you can buy oil at $40 per barrel, hold it for 20 years, and sell it at $1,000 per barrel. The rate of return from this behavior is the solution to (1+r) 20 = 1000 40 .

This equation solves for r = 17.46%, which represents a healthy rate of return on investment. This substitution is part of a general conclusion known as the **Ramsey rule**:The solution to this problem is known as Ramsey pricing, after the discoverer Frank Ramsey (1903–1930). For resources in fixed supply, prices rise at the interest rate. With a resource in fixed supply, owners of the resource will sell at the point maximizing the present value of the resource. Even if they do not, others can buy the resource at the low present value of price point, resell at the high present value, and make money.

The Ramsey rule implies that prices of resources in fixed supply rise at the interest rate. An example of the Ramsey rule in action concerns commodities that are temporarily fixed in supply, such as grains after the harvest. During the period between harvests, these products rise in price on average at the interest rate, where the interest rate includes storage and insurance costs, as well as the cost of funds.

Example: Let time be $t = 0, 1, ...$, and suppose the demand for a resource in fixed supply has constant elasticity: $\mathrm{p(Q)} = \mathrm{aQ} - 1\varepsilon$. Suppose that there is a total stock R of the resource, and the interest rate is fixed at r. What is the price and consumption of the resource at each time?

Solution: Let Qt represent the quantity consumed at time t. Then the arbitrage condition requires

$$\mathrm{a}\ Q0 - 1\varepsilon(1+r)t = p(Q0)(1+r)t = p(Qt) = aQt - 1\varepsilon$$

Thus, $Qt = Q0(1+r) - t\varepsilon$. Finally, the resource constraint implies

$$R = (Q0 + Q1 + Q2 + ...) = Q0(1 + (1+r) - \varepsilon + (1+r) - 2\varepsilon + ...) = Q01 - (1+r) - \varepsilon$$

This solves for the initial consumption Q_0. Consumption in future periods declines geometrically, thanks to the constant elasticity assumption.

Market arbitrage ensures the availability of the resource in the future and drives up the price to ration the good. The world runs out slowly, and the price of a resource in fixed supply rises on average at the interest rate.

Resources like oil and minerals are ostensibly in fixed supply—there is only so much oil, gold, bauxite, or palladium in the earth. Markets, however, behave as if there is an unlimited supply, and with good reason. People are inventive and find substitutes. England's wood shortage of 1651 didn't result in England being cold permanently, nor was England limited to the wood it could grow as a source of heat. Instead, coal was discovered. The shortage of whale oil in the mid-19th century led to the development of oil resources as a replacement. If markets expect that price increases will lead to substitutes, then we rationally should use more today, trusting that technological developments will provide substitutes.Unlike oil and trees, whales were overfished and there was no mechanism for arbitraging them into the future—that is, no mechanism for capturing and saving whales for later use. This problem, known as the tragedy of the commons, results in too much use (Garett Hardin, Science, 1968, Tragedy of the Commons). Trees have also been overcut, most notably on Easter Island. Thus, while some believe that we are running out of oil, most investors are betting that we are not, and that energy will not be very expensive in the future—either because of continued discovery of oil or because of the creation of alternative energy sources. If you disagree, why not invest and take the bet? If you bet on future price increases, that will tend to increase the price today, encouraging conservation today, and increase the supply in the future.

Key Takeaways

- The Ramsey rule holds that, for resources in fixed supply, prices rise at the interest rate.
- With constant elasticity, consumption of a resource in fixed supply declines geometrically.
- Market arbitrage ensures the availability of the resource in the future and drives up the price to ration the good. The world runs out slowly, and the price of a resource in fixed supply rises on average at the interest rate.
- Substitutes mitigate the fixed supply aspect of natural resources; for example, fiber optic cable substitutes for copper.

EXERCISE

1. With an elasticity demand of two, compute the percentage of the resource that is used each year if the interest rate is 10%. If the interest rate falls, what happens to the proportion quantity used?

This page titled 11.4: Resource Extraction is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

11.5: A Time to Harvest 11.5: A Time to HarvestEdit Update titlecancel

Learning Objectives

- How are the prices of renewable resources determined?
- When should trees be harvested?

A tree grows slowly but is renewable, so the analysis of Section 11.4 doesn't help us to understand when it is most profitable to cut down the tree. Consider harvesting for pulp and paper use. In this case, the amount of wood chips is what matters to the profitability of cutting down the tree, and the biomass of the tree provides a direct indication of this. Suppose the biomass sells for a net price *p*, which has the costs of harvesting and replanting deducted from it, and the biomass of the tree is *b*(*t*) when the tree is *t* years old. It simplifies the analysis slightly to use continuous time discounting $\mathrm{e}-\rho\mathrm{t}=(11+r)\mathrm{t},\ \text{where}\ \rho=\log(1+r)$

Consider the policy of cutting down trees when they are *T* years old. This induces a cutting cycle of length *T*. A brand new tree will produce a present value of profits of

$$e-\rho Tpb(T)+e-2\rho Tpb(T)+e-3\rho Tpb(T)+\ldots=e-\rho Tpb(T)1-e-\rho T=pb(T)e\rho T-1.$$

This profit arises because the first cut occurs at time *T*, with discounting $e^{-\rho T}$, and produces a net gain of *pb*(*T*). The process then starts over, with a second tree cut down at time 2*T*, and so on.

Profit maximization gives a first-order condition on the optimal cycle length *T* of

$$0=\mathrm{ddTpb(T)e}\rho\mathrm{T}-1=\mathrm{pb}'(\mathrm{T})\mathrm{e}\rho\mathrm{T}-1-\mathrm{pb(T)}\rho\ \mathrm{e}\ \rho\mathrm{T}(\mathrm{e}\rho\mathrm{T}-1)2$$

This can be rearranged to yield

$$\mathrm{b}'(\mathrm{T})\mathrm{b(T)}=\rho 1-\mathrm{e}-\rho\mathrm{T}$$

The left-hand side of this equation is the growth rate of the tree. The right-hand side is approximately the continuous-time discount factor, at least when *T* is large, as it tends to be for trees, which are usually on a 20- to 80-year cycle, depending on the species. This is the basis for a conclusion: Cut down the tree slightly before it is growing at the interest rate. The higher that interest rates are, the shorter the cycle for which the trees should be cut down.

The pulp and paper use of trees is special, because the tree is going to be ground up into wood chips. What happens when the object is to get boards from the tree, and larger boards sell for more? In particular, it is more profitable to get a 4 × 4 than two 2 × 4s. Doubling the diameter of the tree, which approximately raises the biomass by a factor of six to eight, more than increases the value of the timber by the increase in the biomass.

It turns out that our theory is already capable of handling this case. The only adaptation is a change in the interpretation of the function *b*. Now, rather than representing the biomass, *b*(*t*) must represent the value in boards of a tree that is *t* years old. (The parameter *p* may be set to one.) The only amendment to the rule for cutting down trees is as follows: The most profitable point in time to cut down the tree occurs slightly before the time when the value (in boards) of the tree is growing at the interest rate.

For example, lobsters become more valuable as they grow. The profit-maximizing time to harvest lobsters is governed by the same equation, where *b* (*T*) is the value of a lobster at age *T*. Prohibiting the harvest of lobsters under age *T* is a means of insuring the profit-maximizing capture of lobsters and preventing overfishing.

The implementation of the formula is illustrated in Figure 11.3. The dashed line represents the growth rate $\mathrm{b}'(\mathrm{T})\ \mathrm{b(T)}$, while the solid line represents the discount rate, which was set at 5%. Note that the best time to cut down the trees is when they are approximately 28.7 years old and, at that time, they are growing at 6.5%. Figure 11.3 also illustrates another feature of the optimization—there may be multiple solutions to the optimization problem, and the profit-maximizing solution involves $b'(T)b(T)$ cutting $\rho 1-\mathrm{e}-\rho\mathrm{T}$ from above.

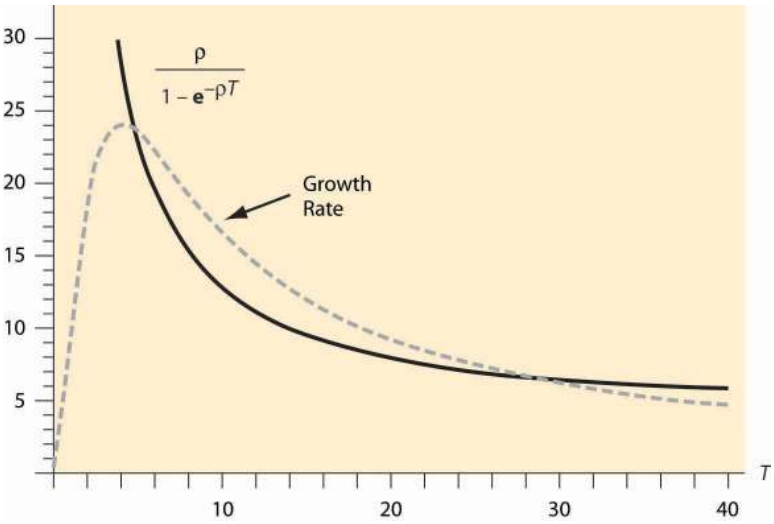


Figure 11.3 Optimal solution for T

The U.S. Department of the Interior is in charge of selling timber rights on federal lands. The department uses the policy of **maximum sustainable yield** to determine the specific time that the tree is cut down. Maximum sustainable yield maximizes the long-run average value of the trees cut down; that is, it maximizes b(T) T .

Maximum sustainable yield is actually a special case of the policies considered here, and arises for a discount factor of 0. It turns out (thanks to a formula known variously as l'Hôpital's or l'Hospital's Rule) that the $\lim \rho \to 0 \rho 1 - \mathrm{e} - \rho \mathrm{T} = 1\mathrm{T}$.

Thus, the rule $\mathrm{b}'(T)\mathrm{b}(T) = \rho 1 - \mathrm{e} - \rho \mathrm{T} \to 1\mathrm{T}$ as $\rho \to 0$, and this is precisely the same rule that arises under maximum sustainable yield.

Thus, the Department of the Interior acts as if the interest rate is zero when it is not. The justification given is that the department is valuing future generations at the same level as current generations—that is, increasing the supply for future generations while slightly harming the current generation of buyers. The major consequence of the department's policy of maximum sustainable yield is to force cutting of timber even when prices are low during recessions.

Key Takeaways

- Renewable resources are harvested slightly earlier than the point at which they grow at the interest rate, because earlier planting of the next generation has value.
- Maximum sustainable yield maximizes the long-run average value of the trees cut down, which is the optimal policy only when the interest rate is zero.

EXERCISES

1. Show that maximum sustainable yield results in cutting down a tree when it is T years old, where T satisfies $b'(T)b(T) = 1T$.
2. Suppose that the growth rate of trees satisfies $b'(T)b(T) = te - t$. Numerically approximate the efficient time to cut down a tree if ρ = 0.1. How does this compare to the solution of maximum sustainable yield?

This page titled 11.5: A Time to Harvest is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

11.6: Collectibles 11.6: CollectiblesEdit Update titlecancel

Learning Objectives

- How are the prices of collectibles determined?
- What is the investment value of collectibles?
- How fast do the prices rise?

Many people purchase durable goods as investments, including Porsche Speedsters (see Figure 11.4), Tiffany lamps, antique telephones, postage stamps and coins, baseball cards, original Barbie dolls, antique credenzas, autographs, original rayon Hawaiian shirts, old postcards, political campaign buttons, old clocks, and even Pez dispensers. How is the value of, say, a 1961 Porsche Speedster or a $500 bill from the Confederacy, which currently sells for over $500, determined?

The theory of resource prices can be adapted to cover these items, which are in fixed supply. There are four major differences that are relevant. First, using the item doesn't consume it: The goods are durable. I can own an "I Like Ike" campaign button for years, and then sell the same button. Second, these items may depreciate. Cars wear out even when they aren't driven, and the brilliant color of Pez dispensers fades. Every time that a standard 27.5-pound gold bar, like the kind in the Fort Knox depository, is moved, approximately $5 in gold wears off the bar. Third, the goods may cost something to store. Fourth, the population grows, and some of the potential buyers are not yet born.



Figure 11.4 The Porsche Speedster

To understand the determinants of the prices of collectibles, it is necessary to create a major simplification to perform the analysis in continuous time. Let t, ranging from zero to infinity, be the continuous-time variable. If the good depreciates at rate δ, and q_0 is the amount available at time 0, the quantity available at time t is q(t)= q 0 e −δt .

For simplicity, assume that there is constant elasticity of demand ε. If g is the population growth rate, the quantity demanded, for any price p, is given by x d (p,t)=a e gt p −ε , for a constant a that represents the demand at time 0. This represents demand for the good for direct use, but neglects the investment value of the good—the fact that the good can be resold for a higher price later. In other words, xd captures the demand for looking at Pez dispensers or driving Porsche Speedsters, but does not incorporate the value of being able to resell these items.

The demand equation can be used to generate the lowest use value to a person owning the good at time t. That marginal use value v arises from the equality of supply and demand:

$$\text{q } 0 \text{ e } -\delta t = q(t) = xd(v,t) = a \text{ e } gtv - \varepsilon$$

or

$$v\varepsilon = aq0e(\delta + g)t$$

Thus, the use value to the marginal owner of the good at time *t* satisfies $\mathrm{v} = (\ \mathrm{a}\ \mathrm{q0})1\varepsilon\ \mathrm{e}\ \delta + \mathrm{g}\varepsilon\mathrm{t}$

An important aspect of this development is that the value to the owner is found without reference to the price of the good. The reason this calculation is possible is that the individuals with high values will own the good, and the number of goods and the values of people are assumptions of the theory. Essentially, we already know that the price will ration the good to the individuals with high values, so computing the lowest value individual who holds a good at time *t* is a straightforward "supply equals demand" calculation. Two factors increase the marginal value to the owner—there are fewer units available because of depreciation, and there are more high-value people demanding them because of population growth. Together, these factors make the marginal use value grow at the rate \(\delta+g \varepsilon \nonumber \] .

Assume that *s* is the cost of storage per unit of time and per unit of the good, so that storing *x* units for a period of length Δ costs *sx*Δ. This is about the simplest possible storage cost technology.

The final assumption we make is that all potential buyers use a common discount rate *r*, so that the discount of money or value received Δ units of time in the future is $e^{-r\Delta}$. It is worth a brief digression to explain why it is sensible to assume a common discount rate, when it is evident that many people have different discount rates. Different discount rates induce gains from trade in borrowing and lending, and create an incentive to have banks. While banking is an interesting topic to study, this section is concerned with collectibles, not banks. If we have different discount factors, then we must also introduce banks, which would complicate the model substantially. Otherwise, we would intermingle the theory of banking and the theory of collectibles. It is probably a good idea to develop a joint theory of banking and collectibles, given the investment potential of collectibles, but it is better to start with the pure theory of either one before developing the joint theory.

Consider a person who values the collectible at *v*. Is it a good thing for this person to own a unit of the good at time *t* ? Let *p* be the function that gives the price across time, so that *p*(*t*) is the price at time *t*. Buying the good at time *t* and then selling what remains (recall that the good depreciates at rate *δ*) at time *t* + Δ gives a net value of $\int 0\Delta e - ru(v-s)du - p(t) + e - r\Delta e - \delta\Delta p(t+\Delta)$.

For the marginal person—that is, the person who is just indifferent to buying or not buying at time *t*—this must be zero at every moment in time, for Δ = 0. If *v* represents the value to a marginal buyer (indifferent to holding or selling) holding the good at time *t*, then this expression should come out to be zero. Thus, dividing by Δ,

$0 = lim\Delta \to 01\Delta \int 0\Delta e - ru(v-s)du - p(t)\Delta + e - (r+\delta)\Delta p(t+\Delta)\Delta = lim\Delta \to 0v - s + p(t+\Delta) - p(t)\Delta - 1 - e - (r+\delta)\Delta\Delta p(t+\Delta) = v - s + p'(t) - (r+\delta)p(t).$

Recall that the marginal value is $v = (aq0)1\varepsilon e\delta + g\varepsilon t$, which gives $p'(t) = (r+\delta)p(t) + s - v = (r+\delta)p(t) + s - (aq0)1\varepsilon e\delta + g\varepsilon t$. The general solution to this differential equation is

$$p(t) = e(r+\delta)t(p(0) + 1 - e - (r+\delta)t(r+\delta)s - (aq0)1\varepsilon 1 - e - (r+\delta-\delta+g\varepsilon)tr + \delta - \delta + g\varepsilon).$$

It turns out that this equation only makes sense if $r + \delta - \delta + g\varepsilon > 0$, for otherwise the present value of the marginal value goes to infinity, so there is no possible finite initial price. Provided that demand is elastic and discounting is larger than growth rates (which is an implication of equilibrium in the credit market), this condition will be met.

What is the initial price? It must be the case that the present value of the price is finite, for otherwise the good would always be a good investment for everyone at time 0, using the "buy and hold for resale" strategy. That is,

$$limt \to \infty e - rtp(t) < \infty.$$

This condition implies that $\lim t \to \infty\ \mathrm{e}\ \delta t(p(0) + 1 - e - (r+\delta)t(r+\delta)s - (aq0)1\varepsilon 1 - e - (r+\delta-\delta+g\varepsilon)tr + \delta - \delta + g\varepsilon) < \infty$, and thus $p(0) + 1(r+\delta)s - (aq0)1\varepsilon 1r + \delta - \delta + g\varepsilon = 0$.

This equation may take on two different forms. First, it may be solvable for a nonnegative price, which happens if

$$p(0) = (aq0)1\varepsilon 1r + \delta - \delta + g\varepsilon - 1(r+\delta)s \geq 0$$

Second, it may require destruction of some of the endowment of the good. Destruction must happen if the quantity of the good q_0 at time 0 satisfies

$$(aq0)1\varepsilon 1r + \delta - \delta + g\varepsilon - 1(r+\delta)s < 0.$$

In this case, there is too much of the good, and an amount must be destroyed to make the initial price zero. Since the initial price is zero, the good is valueless at time zero, and destruction of the good makes sense—at the current quantity, the good is too costly to store for future profits. Enough is destroyed to ensure indifference between holding the good as a collectible and destroying it.

Consider, for example, the $500 Confederate bill shown in Figure 11.5. Many of these bills were destroyed at the end of the U.S. Civil War, when the currency became valueless and was burned as a source of heat. Now, an uncirculated version retails for $900.



Figure 11.5 $500 Confederate States bill

The amount of the good that must be destroyed is such that the initial price is zero. As q_0 is the initial (predestruction) quantity, the amount at time zero after the destruction is the quantity $q(0)$ satisfying

$$0 = p(0) = (aq(0))1\varepsilon 1r + \delta - \delta + g\varepsilon - 1(r+\delta)s.$$

Given this construction, we have that

$$p(0) + 1(r+\delta)s - (aq(0))1\varepsilon 1r + \delta - \delta + g\varepsilon = 0$$

where either $q(0) = q0 and p(0) \geq 0, or q(0) < q0 and p(0) = 0.$

Destruction of a portion of the stock of a collectible, followed by price increases, is actually a quite common phenomenon. In particular, consider the “Model 500” telephone by Western Electric illustrated in Figure 11.6. This ubiquitous classic phone was retired as the United States switched to tone dialing and push-button phones in the 1970s, and millions of phones—perhaps over 100 million—wound up in landfills. Now the phone is a collectible, and rotary phone enthusiasts work to keep them operational.



Figure 11.6 Western Electric Model 500 telephone

The solution for $p(0)$ dramatically simplifies the expression for $p(t)$:

$$p(t) = e(r+\delta)t(p(0) + 1 - e - (r+\delta)t(r+\delta)s - (aq(0))1\varepsilon 1 - e - (r+\delta-\delta+g\varepsilon)tr + \delta - \delta + g\varepsilon)$$
$$= e(r+\delta)t(-e - (r+\delta)t(r+\delta)s + (aq(0))1\varepsilon e - (r+\delta-\delta+g\varepsilon)tr + \delta - \delta + g\varepsilon) = (aq(0))1\varepsilon e\delta + g\varepsilon tr + \delta - \delta + g\varepsilon - sr + \delta$$

This formula enables one to compare different collectibles. The first insight is that storage costs enter linearly into prices, so that growth rates are approximately unaffected by storage costs. The fact that gold is easy to store—while stamps and art require control of humidity and temperature in order to preserve value, and are hence more expensive to store—affects the level of prices but not the growth rate. However, depreciation and the growth of population affect the growth rate, and they do so in combination with the demand elasticity. With more elastic demand, prices grow more slowly and start at a lower level.

Key Takeaways

- The price of collectibles includes two distinct sources of value—use value and investment value. The relevant use value is that of the marginal user, a value that rises as the quantity falls or the population grows.
- The use value to the marginal owner is found without reference to the price of the good. It grows at the destruction rate plus the population growth rate, all divided by the demand elasticity.
- The investment value net of storage must equal the interest rate.

- If storage costs are high, equilibrium pricing may first involve destruction of some quantity of units.
- Storage costs enter linearly into prices, so that growth rates are approximately unaffected by storage costs. The fact that gold is easy to store—while stamps and art require control of humidity and temperature in order to preserve value, and are hence more expensive to store—affects the level of prices but not the growth rate. However, depreciation and the growth of population affect the growth rate, and they do so in combination with the demand elasticity. With more elastic demand, prices grow more slowly and start at a lower level.

This page titled 11.6: Collectibles is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

11.7: Summer Wheat 11.7: Summer WheatEdit Update titlecancel

Learning Objectives

- How are the prices of storable goods—apples, potatoes, or wheat—determined?

Typically, wheat harvested in the fall has to last until the following harvest. How should prices evolve over the season? If I know that I need wheat in January, should I buy it at harvest time and store it myself, or wait and buy it in January? We can use a theory analogous to the theory of collectibles developed in Section 11.6 to determine the evolution of prices for commodities like wheat, corn, orange juice, and canola oil.

Unlike collectibles, buyers need not hold commodities for their personal use, since there is no value in admiring the wheat in your home. Let *p*(*t*) be the price at time *t*, and suppose that the year has length *T*. Generally there is a substantial amount of uncertainty regarding the size of wheat harvests, and most countries maintain an excess inventory as a precaution. However, if the harvest were not uncertain, there would be no need for a precautionary holding. Instead, we would consume the entire harvest over the course of a year, at which point the new harvest would come in. It is this such model that is investigated in this section.

Let *δ* represent the depreciation rate (which, for wheat, includes the quantity eaten by rodents), and let *s* be the storage cost. Buying at time *t* and reselling at *t* + Δ should be a break-even proposition. If one purchases at time *t*, it costs *p* (*t*) to buy the good. Reselling at *t* + Δ, the storage cost is about *s*Δ. (This is not the precisely relevant cost; but rather it is the present value of the storage cost, and hence the restriction to small values of Δ.) The good depreciates to only have e −δΔ left to sell, and discounting reduces the value of that amount by the factor e −rΔ . For this to be a break-even proposition, for small Δ,

$$0 = e - r\Delta e - \delta\Delta p(t+\Delta) - s\Delta - p(t)$$

or

$$p(t+\Delta) - p(t)\Delta = 1 - e - (r+\delta)\Delta\Delta p(t+\Delta) + s$$

taking the limit as $\Delta \to 0, \mathrm{p}'(\mathrm{t}) = (\mathrm{r}+\delta)\mathrm{p}(\mathrm{t}) + \mathrm{s}$

This arbitrage condition ensures that it is a break-even proposition to invest in the good; the profits from the price appreciation are exactly balanced by depreciation, interest, and storage costs. We can solve the differential equation to obtain

$$p(t) = e(r+\delta)t(p(0) + 1 - e - (r+\delta)tr + \delta s) = e(r+\delta)tp(0) + e(r+\delta)t - 1r + \delta s$$

The unknown is *p* (0). The constraint on *p* (0), however, is like the resource extraction problem—*p* (0) is determined by the need to use up the harvest over the course of the year.

Suppose demand has constant elasticity *ε*. Then the quantity used comes in the form $x(t) = \mathrm{ap}(t) - \varepsilon$. Let *z*(*t*) represent the stock at time *t*. Then the equation for the evolution of the stock is $z'(t) = -x(t) - \delta z(t)$. This equation is obtained by noting that the flow out of stock is composed of two elements: depreciation, *δz*, and consumption, *x*. The stock evolution equation solves for

$$z(t) = e - \delta t(q(0) - \int 0te\delta ux(u)du).$$

Thus, the quantity of wheat is consumed exactly if

$$\int 0\ \mathrm{T}\ \mathrm{e}\ \delta ux(u)du = q(0)$$

But this equation determines the initial price through

$$q(0) = \int 0Te\delta ux(u)du = \int 0Te\delta uap(u) - \varepsilon du = \int 0Te\delta ua(e(r+\delta)up(0) + e(r+\delta)u - 1r + \delta s) - \varepsilon du.$$

This equation doesn't lead to a closed form for *p*(0) but is readily estimated, which provides a practical means of computing expected prices for commodities in temporarily fixed supply.

Figure 11.7 Prices over a cycle for seasonal commodities

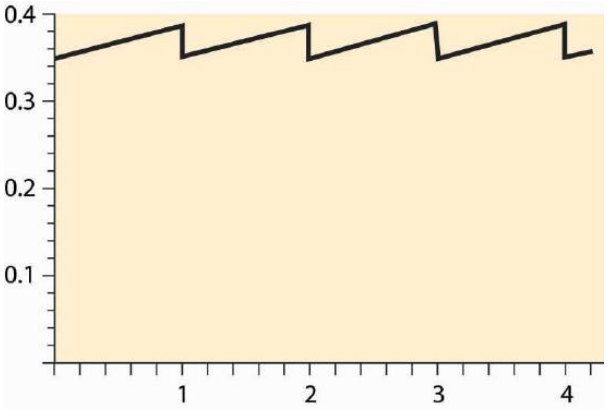


Figure 11.7. The increasing portion is actually an exponential, but of such a small degree that it looks linear. When the new harvest comes in, prices drop abruptly as the inventory grows dramatically, and the same pattern is repeated.

Figure 11.8 Log of price of gold over time

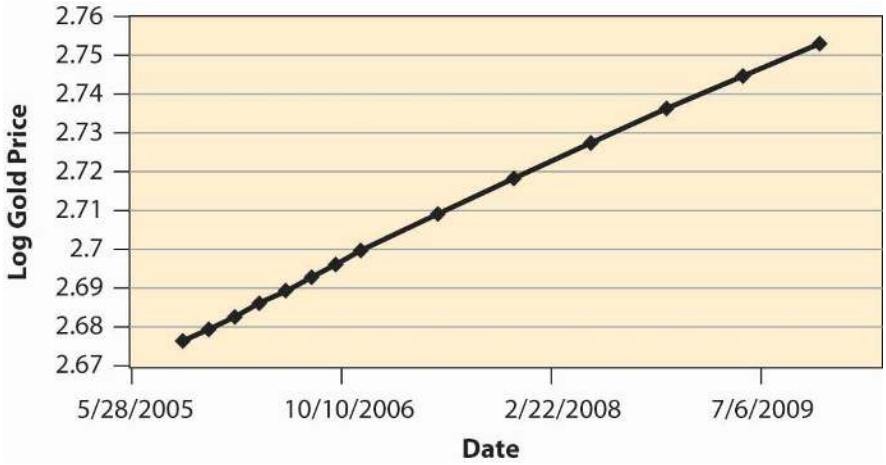


Figure 11.8 shows the log of the future price of gold over time. The relevant data come from a futures market that establishes, at one moment in time, the price of gold for future delivery, and thus represents today's estimate of the future price of gold. These data, then, represent the expected future price at a particular moment in time (the afternoon of October 11, 2005), and thus correspond to the prices in the theory, since perceived risks are fixed. (Usually, in the real world, risk plays a salient role.) We can observe that prices are approximately an exponential, because the log of prices is approximately linear. However, the estimate of $r + \delta$ is surprisingly low, at an annual level of less than 0.03, or 3% for both discounting and depreciation. Depreciation of gold is low, but this still represents a very low interest rate.

Key Takeaways

- There is a seasonal pattern to goods that are produced periodically. The price equation produces a "sawtooth" pattern. The increasing portion is an exponential.
- Gold prices show evidence of exponential growth predicted by the theory.

EXERCISE

1. Consider a market for a commodity that can be stored with zero cost from winter to summer, but cannot be stored from summer to winter. The winter demand and supply are $Qwd = 50 - 2Pw$ and $Qws = 3Pw$, and the summer demand and supply are $Qsd = 100 - 3Ps$ and $Qss = Ps$. Compute Pw, Ps, Qw, and Qs, and the amount of hoarding from winter to summer. (Set discounting to zero.)

This page titled 11.7: Summer Wheat is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

12: Consumer Theory

12.1: Utility Maximization
12.2: Budget or Feasible Set
12.3: Isoquants
12.4: Examples
12.5: Substitution Effects
12.6: Income Effects
12.7: Mathematical Cleanup

This page titled 12: Consumer Theory is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

12.1: Utility Maximization 12.1: Utility MaximizationEdit Update titlecancel

Learning Objectives

- How do economists model consumer choice?

Economists use the term utility in a peculiar and idiosyncratic way. **Utility** refers not to usefulness but to the flow of pleasure or happiness that a person enjoys—some measure of the satisfaction a person experiences. Usefulness might contribute to utility, but so does style, fashion, or even whimsy.

The term utility is unfortunate, not just because it suggests usefulness but because it makes the economic approach to behavior appear more limited than it actually is. We will make very few assumptions about the form of utility that a consumer might have. That is, we will attempt to avoid making value judgments about the preferences a consumer holds—whether he or she likes to smoke cigarettes or eat only carrots, watch Arnold Schwarzenegger movies, or spend time with a hula-hoop. Consumers like whatever it is that they like; the economic assumption is that they attempt to obtain the goods that they enjoy. It is the consequences of the pursuit of happiness that comprise the core of consumer theory.

In this chapter, we will focus on two goods. In many cases, the generalization to an arbitrary number of goods is straightforward. Moreover, in most applications it won't matter because we can view one of the goods as a "composite good," reflecting consumption of a bunch of other goods.Thus, for example, savings for future consumption, providing for descendants, or giving to your alma mater are all examples of consumption. Our consumer will, in the end, always spend all of his or her income, although this happens because we adopt a very broad notion of spending. In particular, savings are "future spending."

As a starting point, suppose there are two goods, *X* and *Y*. To distinguish the quantity of the good from the good itself, we'll use capital letters to indicate the good, and lowercase letters to indicate the quantity of that good that is consumed. If *X* is rutabagas, a consumer who ate three of them would have $x = 3$. How can we represent preferences for this consumer? To fix ideas, suppose the consumer is both hungry and thirsty, and the goods are pizza and beer. The consumer would like more of both, reflected in greater pleasure for greater consumption. Items that one might consume are generally known as "bundles," as in bundles of goods and services, and less frequently as "tuples," a short form for the "*n*-tuple," meaning a list of *n* quantities. Since we will focus on two goods, both of these terms are strained in the application—a bundle because a bundle of two things isn't much of a bundle, and a tuple because what we have here is a "two-tuple," also known as a pair. But part of the job of studying economics is to learn the language of economics, so bundles it is.

One might naturally consider measuring utility on some kind of physical basis (production of dopamine in the brain, for example) but it turns out that the actual quantities of utility don't matter for the theory we develop. What matters is whether a bundle produces more than another, less, or the same. Let $u(x, y)$ represent the utility that a consumer gets from consuming *x* units of beer and *y* units of pizza. The function *u* guides the consumer's choice in the sense that, if the consumer can choose either (x_1, y_1) or (x_2, y_2), we expect him to choose (x_1, y_1) if $u(x_1, y_1) > u(x_2, y_2)$ But notice that a doubling of *u* would lead to the same choices because

$$u(x_1, y_1) > u(x_2, y_2) \text{ if and only if } 2u(x_1, y_1) > 2u(x_2, y_2)$$

Thus, doubling the utility doesn't change the preferences of the consumer. But the situation is more extreme than this. Even exponentiating the utility doesn't change the consumer's preferences because $u(x_1, y_1) > u(x_2, y_2)$ if and only if e $u(x_1, y_1) > eu(x_2, y_2)$

In other words, there are no natural units for utility, at least until such time as we are able to measure pleasure in the brain.

It is possible to develop the theory of consumer choice without supposing that a utility function exists at all. However, it is expedient to begin with utility in order to simplify the analysis for introductory purposes.

Key Takeaways

- Consumer theory is to demand as producer theory is to supply.
- Consumer theory is based on the premise that we can infer what people like from the choices they make.
- Utility refers not to usefulness but to the flow of pleasure or happiness that a person enjoys—some measure of the satisfaction a person experiences.
- There are no natural units for utility; any increasing transformation is acceptable.

This page titled 12.1: Utility Maximization is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

12.2: Budget or Feasible Set 12.2: Budget or Feasible SetEdit Update titlecancel

Learning Objectives

- How does income limit choice?

Suppose that a consumer has a fixed amount of money to spend, M. There are two goods X and Y, with associated prices p_X and p_Y. The feasible choices that the consumer can make satisfy p X x+ p Y y≤M. In addition, we will focus on consumption and rule out negative consumption, so $x \geq 0$ and $y \geq 0$. This gives a **budget set** or feasible set, as illustrated in Figure 12.1. The budget set is the set of goods a consumer can afford to purchase.

The **budget line** is the boundary of the budget set, and it consists of the goods that just exhaust the consumer's budget

Figure 12.1 Budget set

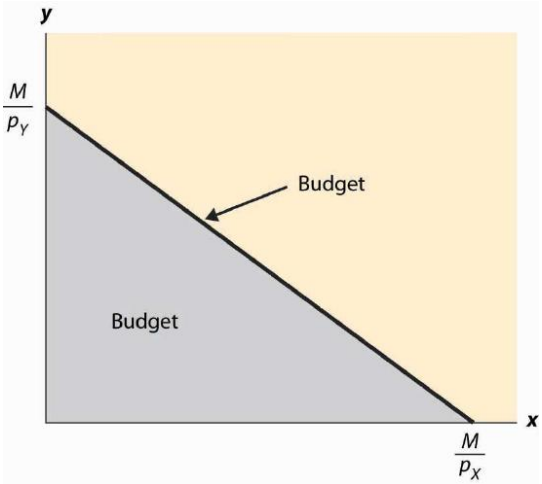


Figure 12.1, the feasible set of purchases that satisfies the budget constraint is illustrated with shading. If the consumer spends all of her money on X, she can consume the quantity x = M p X . Similarly, if she spends all of her money on Y, she consumes M p Y units of Y. The straight line between them, known as the budget line, represents the most of the goods that she can consume. The slope of the budget line is − p X p Y .

An increase in the price of one good pivots or rotates the budget line. Thus, if the price of X increases, the endpoint M p Y remains the same, but M p X falls. This is illustrated in Figure 12.2.

Figure 12.2 Effect of an increase in price on the budget

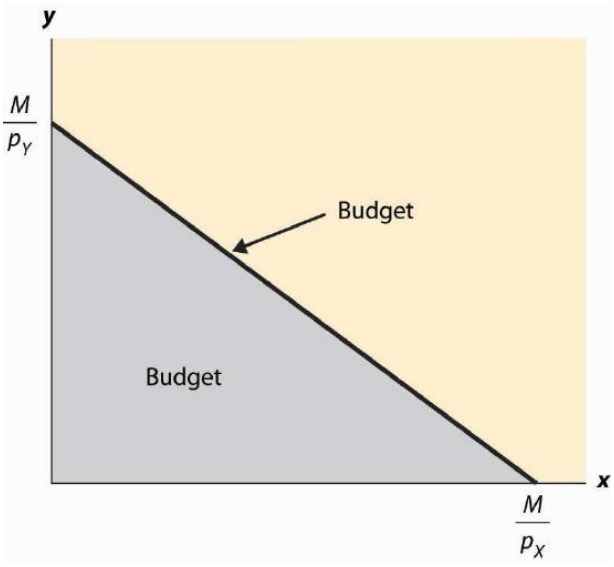


Figure 12.3.

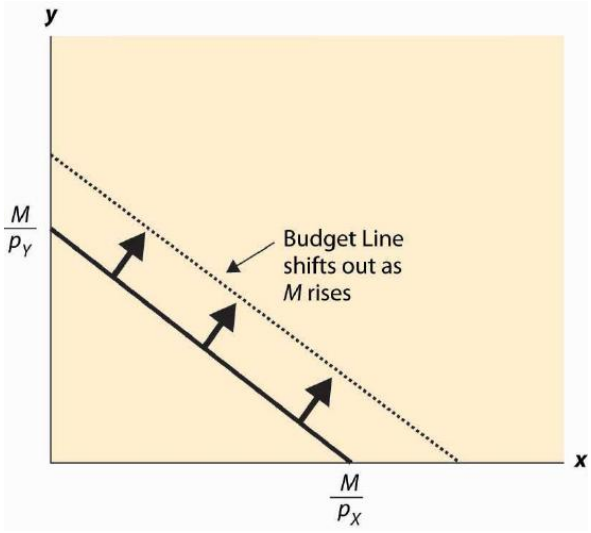


Figure 12.3 An increase in income

An increase in both prices by the same proportional factor has an effect identical to a decrease in income. Thus, one of the three financial values—the two prices and income—is redundant. That is, we can trace out all of the possible budget lines with any two of the three parameters. This can prove useful. We can arbitrarily set pX to be the number one without affecting the generality of the analysis. When setting a price to one, that related good is called the **numeraire**, and essentially all prices are denominated with respect to that one good.

A real-world example of a numeraire occurred when the currency used was based on gold, so that the prices of other goods were denominated in terms of the value of gold.

Money is not necessarily the only constraint on the consumption of goods that a consumer faces. Time can be equally important. One can own all of the compact disks in the world, but they are useless if one doesn't actually have time to listen to them. Indeed, when we consider the supply of labor, time will be a major issue—supplying labor (working) uses up time that could be used to consume goods. In this case, there will be two kinds of budget constraints—a financial one and a temporal one. At a fixed wage, time and money translate directly into one another, and the existence of the time constraint won't present significant challenges to the theory. The conventional way to handle the time constraint is to use, as a baseline, working "full out," and then to view leisure as a good that is purchased at a price equal to the wage. Thus, if you earn $20 an hour, we would set your budget at $480 a day, reflecting 24 hours of work; but we would then permit you to buy leisure time, during which eating, sleeping, brushing your teeth, and every other nonwork activity could be accomplished at a price equal to $20 per hour.

Key Takeaways

- The budget set or feasible set is the set of goods that the consumer can afford to purchase.
- The budget line is the pair of goods that exactly spend the budget. The budget line shifts out when income rises and pivots when the price of one good changes.
- Increasing prices and income by the same multiplicative factor leaves the feasible set unchanged.

EXERCISES

1. Graph the budget line for apples and oranges with prices of $2 and $3, respectively, and $60 to spend. Now increase the price of apples from $2 to $4, and draw the budget line.
2. Suppose that apples cost $1 each. Water can be purchased for 50 cents per gallon up to 20,000 gallons, and 10 cents per gallon for each gallon beyond 20,000 gallons. Draw the budget constraint for a consumer who spends $200 on apples and water.
3. Graph the budget line for apples and oranges with prices of $2 and $3, respectively, and $60 to spend. Now increase the expenditure to $90, and draw the budget line.

This page titled 12.2: Budget or Feasible Set is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

12.3: Isoquants 12.3: IsoquantsEdit Update titlecancel

Learning Objectives

- What is an isoquant?
- Why does it help to analyze consumer choice?

With two goods, we can graphically represent utility by considering the contour map of utility. Utility contours are known as isoquants, meaning "equal quantity," and are also known as indifference curves, since the consumer is indifferent between points on the line. In other words an **indifference curve**, also known as an iso-utility curve, is the set of goods that produce equal utility.

We have encountered this idea already in the description of production functions, where the curves represented input mixes that produced a given output. The only difference here is that the output being produced is consumer "utility" instead of a single good or service.

Figure 12.4 Utility isoquants

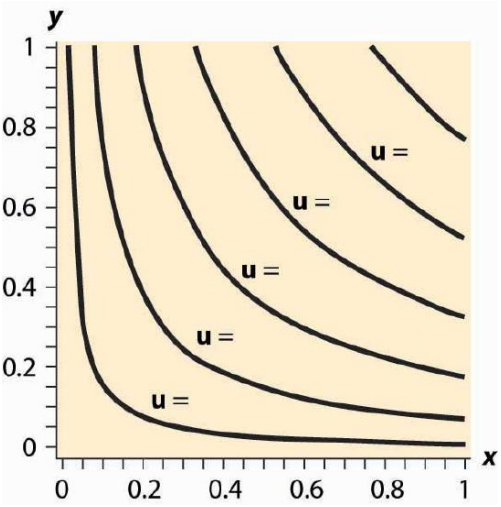


Figure 12.4 provides an illustration of isoquants, or indifference curves. Each curve represents one level of utility. Higher utilities occur to the northeast, farther away from the origin. As with production isoquants, the slope of the indifference curves has the interpretation of the trade-off between the two goods. The amount of *Y* that the consumer is willing to give up, in order to obtain an extra bit of *X*, is the slope of the indifference curve. Formally, the equation u(x,y)= u o defines an indifference curve for the reference utility u_0. Differentiating in such a way as to preserve the equality, we obtain the slope of the indifference curve:

$$\partial u \partial x dx + \partial u \partial y dy = 0$$

$$\mathrm{dydx|u = u0} = -\partial \mathrm{u} \partial \mathrm{x} \partial \mathrm{u} \partial \mathrm{y}$$

This slope is known as the marginal rate of substitution and reflects the trade-off, from the consumer's perspective, between the goods. That is to say, the marginal rate of substitution (of *Y* for *X*) is the amount of *Y* that the consumer is willing to lose in order to obtain an extra unit of *X*.

An important assumption concerning isoquants is reflected in the figure: "Midpoints are preferred to extreme points." Suppose that the consumer is indifferent between (x_1, y_1) and (x_2, y_2); that is, u (x_1, y_1) = u (x_2, y_2). Then we can say that preferences are convex if any point on the line segment connecting (x_1, y_1) and (x_2, y_2) is at least as good as the extremes. Formally, a point on the line segment connecting (x_1, y_1) and (x_2, y_2) comes in the form $(\alpha x_1 + (1-\alpha)x_2, \alpha y_1 + (1-\alpha)y_2)$ for α between zero and one. This is also known as a "convex combination" between the two points. When α is zero, the segment starts at (x_2, y_2) and proceeds in a linear fashion to (x_1, y_1) at α = 1. Preferences are convex if, for any α between 0 and 1, $u(x_1, y_1) = u(x_2, y_2)$ implies $u(\alpha x_1 + (1-\alpha)x_2, (y_1 + (1-\alpha)\ y_2) \geq u(x_1, y_1)$

This property is illustrated in Figure 12.5. The line segment that connects two points on the indifference curve lies to the northeast of the indifference curve, which means that the line segment involves strictly more consumption of both goods than some points on the indifference curve. In other words, it is preferred to the indifference curve. Convex preferences mean that a consumer prefers a mix to any two equally valuable extremes. Thus, if the consumer likes black coffee and also likes drinking milk, then the consumer

prefers some of each—not necessarily mixed—to only drinking coffee or only drinking milk. This sounds more reasonable if you think of the consumer's choices on a monthly basis. If you like drinking 60 cups of coffee and no milk per month as much as you like drinking 30 glasses of milk and no coffee, convex preferences entail preferring 30 cups of coffee and 15 glasses of milk to either extreme.

How does a consumer choose which bundle to select? The consumer is faced with the problem of maximizing u (x, y) subject to p X x+ p Y y≤M.We can derive the solution to the consumer's problem as follows. First, "solve" the budget constraint p X x+ p Y y≤M for y, to obtain y≤ M− p X x p Y . If Y is a good, this constraint will be satisfied with equality, and all of the money will be spent. Thus, we can write the consumer's utility as

$$u(x, M - pX \times pY)$$

The first-order condition for this problem, maximizing it over x, has

$$0 = \mathrm{ddxu(x, M - pX \times pY) = \partial u \partial x - pXpY \partial u \partial y}$$

.

This can be rearranged to obtain the marginal rate of substitution (MRS):

\partial is only supported in math mode

The **marginal rate of substitution (MRS)** is the extra amount of one good needed to make up for a decrease in another good, staying on an indifference curve

The first-order condition requires that the slope of the indifference curve equals the slope of the budget line; that is, there is a tangency between the indifference curve and the budget line. This is illustrated in Figure 12.6. Three indifference curves are drawn, two of which intersect the budget line but are not tangent. At these intersections, it is possible to increase utility by moving "toward the center," until the highest of the three indifference curves is reached. At this point, further increases in utility are not feasible, because there is no intersection between the set of bundles that produce a strictly higher utility and the budget set. Thus, the large black dot is the bundle that produces the highest utility for the consumer.

It will later prove useful to also state the second-order condition, although we won't use this condition now:

$$0 \geq \mathrm{d2(dx)2u(x, M - pX \times pY) = \partial 2u(\partial x)2 - pXpY\partial 2u\partial x\partial y + (pXpY)2\partial 2u(\partial y)2}$$

Note that the vector $(u1, u2) = (\partial u \partial x, \partial u \partial y)$ is the gradient of u, and the gradient points in the direction of steepest ascent of the function u. Second, the equation that characterizes the optimum,

$$0 = \mathrm{p \times \partial u \partial y - pY \partial u \partial x = (\partial u \partial x, \partial u \partial y) \cdot (-pY, pX)}$$

,

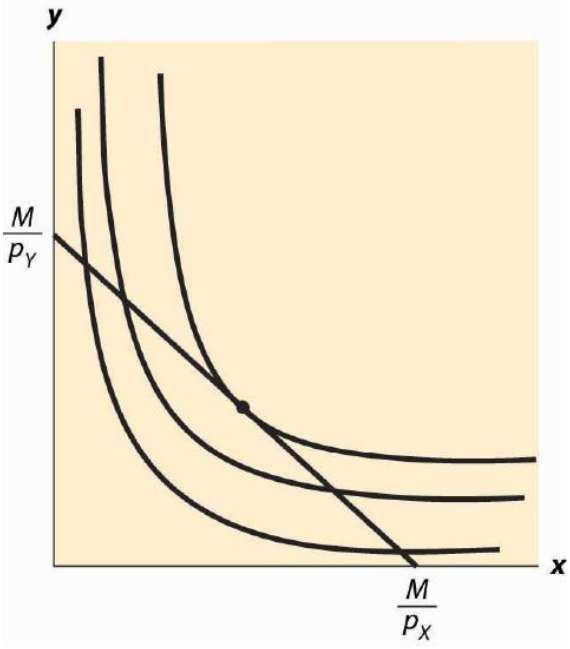


Figure 12.6 Graphical utility maximization

where • is the "dot product" that multiplies the components of vectors and then adds them, says that the vectors (u_1, u_2) and (–pY, pX) are perpendicular and, hence, that the rate of steepest ascent of the utility function is perpendicular to the budget line.

When does this tangency approach fail to solve the consumer's problem? There are three ways that it can fail. First, the utility might not be differentiable. We will set aside this kind of failure with the remark that fixing points of nondifferentiability is mathematically challenging but doesn't lead to significant alterations in the theory. The second failure is that a tangency doesn't maximize utility. Figure 12.7 illustrates this case. Here there is a tangency, but it doesn't maximize utility. In Figure 12.7 , the dotted indifference curve maximizes utility given the budget constraint (straight line). This is exactly the kind of failure that is ruled out by convex preferences. In Figure 12.7, preferences are not convex because, if we connect two points on the indifference curves and look at a convex combination, we get something less preferred, with lower utility—not more preferred as convex preferences would require.

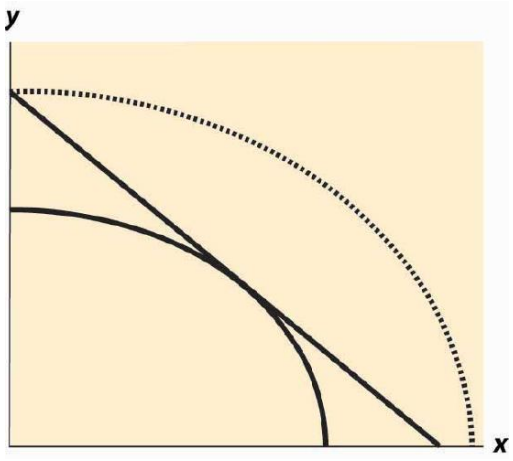


Figure 12.7 "Concave" preferences: Prefer boundaries

The third failure is more fundamental: The derivative might fail to be zero because we've hit the boundary of x = 0 or y = 0. This is a fundamental problem because, in fact, there are many goods that we do buy zero of, so zeros for some goods are not uncommon solutions to the problem of maximizing utility. We will take this problem up in a separate section, but we already have a major tool to deal with it: convex preferences. As we shall see, convex preferences ensure that the consumer's maximization problem is "well behaved."

Key Takeaways

- Isoquants, meaning “equal quantity,” are also known as indifference curves and represent sets of points holding utility constant. They are analogous to production isoquants.
- Preferences are said to be convex if any point on the line segment connecting a pair of points with equal utility is preferred to the endpoints. This means that whenever the consumer is indifferent between two points, he or she prefers a mix of the two.
- The first-order conditions for the maximizing utility involve equating the marginal rate of substitution and the price ratio.
- At the maximum, the rate of steepest ascent of the utility function is perpendicular to the budget line.
- There are two main ways that the first-order conditions fail to characterize the optimum: the consumer doesn’t have convex preferences, or the optimum involves a zero consumption of one or more goods.

This page titled 12.3: Isoquants is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

12.4: Examples 12.4: ExamplesEdit Update titlecancel

Learning Objectives

- Are there any convenient functional forms for analyzing consumer choice?

The Cobb-Douglas utility function comes in the form u(x,y)= x α y 1−α . Since utility is zero if either of the goods is zero, we see that a consumer with Cobb-Douglas preferences will always buy some of each good. The marginal rate of substitution for Cobb-Douglas utility is

$$-dydx|u = u0 = \partial u\partial x\partial u\partial y = ay(1-a)x$$

Thus, the consumer's utility maximization problem yields

$$pXpY = -dydx|u = u0 = \partial u\partial x\partial u\partial y = ay(1-a)x$$

Thus, using the budget constraint, $(1-a) \times pX = aypY = a(M - xpX)$ This yields

$$\mathrm{x} = \mathrm{aMpX}, \quad \mathrm{y} = (1-\mathrm{a})\mathrm{MpY}$$

The Cobb-Douglas utility results in constant expenditure shares. No matter what the price of X or Y, the expenditure xpX on X is αM. Similarly, the expenditure on Y is (1 – α)M. This makes the Cobb-Douglas utility very useful for computing examples and homework exercises.

When two goods are perfect complements, they are consumed proportionately. The utility that gives rise to perfect complements is in the form $u(x,y) = \min\{x, \beta y\}$ for some constant β (the Greek letter "beta"). First observe that, with perfect complements, consumers will buy in such a way that x = βy. The reason is that, if x > βy, some expenditure on x is a waste since it brings in no additional utility; and the consumer gets higher utility by decreasing x and increasing y. This lets us define a "composite good" that involves buying some amount y of Y and also buying βy of X. The price of this composite commodity is $\beta\mathrm{pX} + \mathrm{pY}$, and it produces utility $u = M\beta pX + pY$. In this way, perfect complements boil down to a single good problem.

If the only two goods available in the world were pizza and beer, it is likely that **satiation**—the point at which increased consumption does not increase utility—would set in at some point. How many pizzas can you eat per month? How much beer can you drink? (Don't answer that.)

Figure 12.8 Isoquants for a bliss point

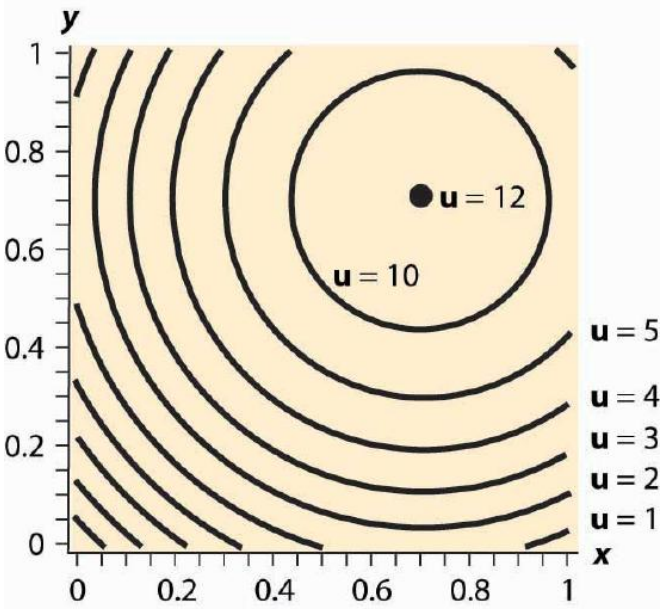


Figure 12.8. Near the origin, the isoquants behave as before. However, as one gets full of pizza and beer, a point of maximum value is reached, illustrated by a large black dot. What does satiation mean for the theory? First, if the bliss point isn't within reach, the theory behaves as before. With a bliss point within reach, consumption will stop at the bliss point. A feasible bliss point entails having a zero value of money. There may be people with a zero value of money, but even very wealthy people, who reach satiation in goods that they personally consume, often like to do other things with the wealth and appear not to have reached satiation overall.

Key Takeaways

- The Cobb-Douglas utility results in constant expenditure shares.
- When two goods are perfect complements, they are consumed proportionately. Perfect complements boil down to a single good problem.
- A bliss point, or satiation, is a point at which further increases in consumption reduce utility.

EXERCISES

1. Consider a consumer with utility $u(x, y) = xy$. If the consumer has \$100 to spend, and the price of X is \$5 and the price of Y is \$2, graph the budget line; and then find the point that maximizes the consumer's utility given the budget. Draw the utility isoquant through this point. What are the expenditure shares?
2. Consider a consumer with utility $u(x, y) = xy$. Calculate the slope of the isoquant directly by solving u(x,y)= u 0 for y as a function of x and the utility level u_0. What is the slope $-dydx|u = u0$? Verify that it satisfies the formula given above.
3. Consider a consumer with utility $u(x, y) = (xy)2$. Calculate the slope of the isoquant directly by solving u(x,y)= u 0 for y as a function of x and the utility level u_0. What is the slope − dy dx | u= u 0 ? Verify that the result is the same as in the previous exercise. Why is it the same?
4. The case of perfect substitutes arises when all that matters to the consumer is the sum of the products—for example, red shirts and green shirts for a colorblind consumer. In this case, $u(x, y) = x + y$. Graph the isoquants for perfect substitutes. Show that the consumer maximizes utility by spending his or her entire income on whichever product is cheaper.
5. Suppose u(x,y)= x α + y α for α < 1. Show that $x = MpX(1 + (pYpX)a)$ and $y = MpY(1 + (pXpY)a)$.
6. Suppose that one consumer has the utility function u (which is always a positive number), and a second consumer has utility w. Suppose, in addition, that for any x, y, $w(x, y) = (u(x, y))^2$; that is, the second person's utility is the square of the first person's. Show that these consumers make the same choices—that is, $u \geq x_a, y_a) \geq u\,(x_b, y_b)$ if and only if $w\,(x_a, y_a) \geq w\,(x_b, y_b)$

This page titled 12.4: Examples is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

12.5: Substitution Effects 12.5: Substitution EffectsEdit Update titlecancel

Learning Objectives

- When prices change, how do consumers change their behavior?

It would be a simpler world if an increase in the price of a good always entailed buying less of it. Alas, it isn't so, as Figure 12.9 illustrates. In this figure, an increase in the price of Y causes the budget line to pivot around the intersection on the x-axis, since the amount of X that can be purchased hasn't changed. In this case, the quantity y of Y demanded rises.

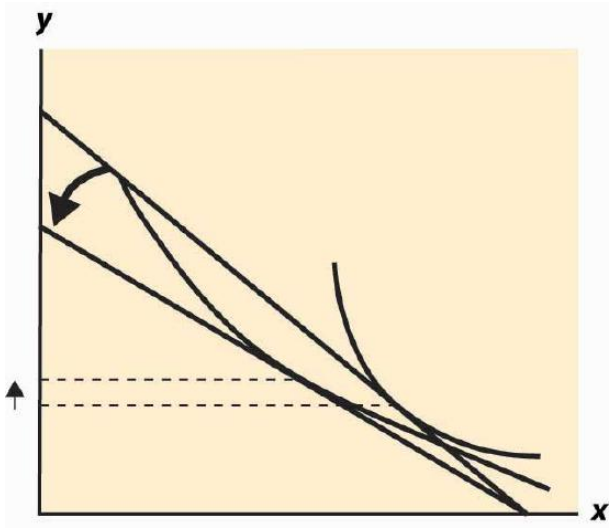


Figure 12.9 Substitution with an increase in price

At first glance, this increase in the consumption of a good in response to a price increase sounds implausible, but there are examples where it makes sense. The primary example is leisure. As wages rise, the cost of leisure (forgone wages) rises. But as people feel wealthier, they choose to work fewer hours. The other examples given, which are hotly debated in the "tempest in a teapot" kind of way, involve people subsisting on a good like potatoes but occasionally buying meat. When the price of potatoes rises, they can no longer afford meat and buy even more potatoes than before.

Thus, the logical starting point on substitution—what happens to the demand for a good when the price of that good increases—does not lead to a useful theory. As a result, economists have devised an alternative approach based on the following logic. An increase in the price of a good is really a composition of two effects: an increase in the relative price of the good and a decrease in the purchasing power of money. As a result, it is useful to examine these two effects separately. The **substitution effect** considers the change in the relative price, with a sufficient change in income to keep the consumer on the same utility isoquant.Some authors instead change the income enough to make the old bundle affordable. This approach has the virtue of being readily computed, but the disadvantage is that the substitution effect winds up increasing the utility of the consumer. Overall the present approach is more economical for most purposes. The income effect changes only income.

Figure 12.10 Substitution effect

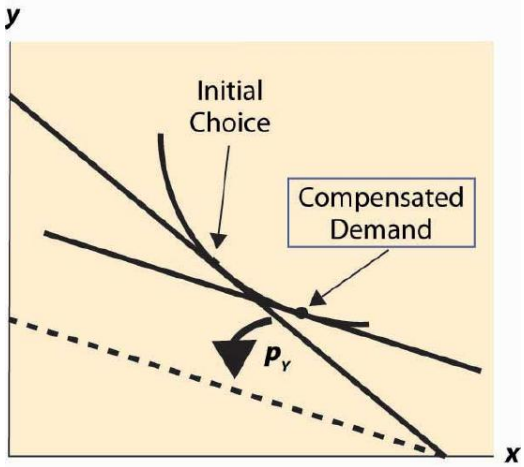


Figure 12.10. The starting point is the tangency between the isoquant and the budget line, denoted with a diamond shape and labeled "Initial Choice." The price of Y rises, pivoting the budget line inward. The new budget line is illustrated with a heavy, dashed line. To find the substitution effect, increase income from the dashed line until the original isoquant is reached. Increases in income shift the budget line out in a fashion parallel to the original. We reach the original isoquant at a point labeled with a small circle, a point sometimes called the **compensated demand** because we have compensated the consumer for the price increase by increasing income just enough to leave her unharmed, on the same isoquant. The substitution effect is just the difference between these points—the substitution in response to the price change, holding constant the utility of the consumer.

We can readily see that the substitution effect of a price increase in Y is to decrease the consumption of Y and increase the consumption of X.To construct a formal proof, first show that if p_Y rises and y rises, holding utility constant, the initial choice prior to the price increase is feasible after the price increase. Use this to conclude that, after the price increase, it is possible to have strictly more of both goods, contradicting the hypothesis that utility was held constant. The **income effect** is the change in consumption resulting from the change in income. The effect of any change in price can be decomposed into the substitution effect, which holds utility constant while changing and the income effect, which adjusts for the loss of purchasing power arising from the price increase.

Example (Cobb-Douglas): Recall that the Cobb-Douglas utility comes in the form u(x,y)= x α y 1−α . Solving for x, y we obtain

$x = aMpX, \quad y = (1-a)MpY$

and

$$u(x,y) = aa(1-a)1-aMpXapY1-a$$

Thus, consider a multiplicative increase Δ in p_Y that is, multiplying pY by Δ > 1. In order to leave the utility constant, M must rise by $\Delta^{1-\alpha}$. Thus, x rises by the factor $\Delta^{1-\alpha}$ and y falls by the factor $\Delta^{-\alpha} < 1$. This is the substitution effect.

What is the substitution effect of a small change in the price pY for any given utility function, not necessarily Cobb-Douglas? To address this question, it is helpful to introduce some notation. We will subscript the utility to indicate partial derivative; that is,

$u1 = \partial u \partial x$, u2 = ∂u∂y.

Note that, by the definition of the substitution effect, we are holding the utility constant, so u(x, y) is being held constant. This means, locally, that 0=du= u 1 dx+ u 2 dy. Writing dx for an unknown infinitesimal change in x can be put on a formal basis. The easiest way to do so is to think of dx as representing the derivative of x with respect to a parameter, which will be p_Y.

In addition, we have $\mathrm{M} = \mathrm{pXx} + \mathrm{pYy}, \text{ so } \mathrm{dM} = \mathrm{pXdx} + \mathrm{pYdy} + \mathrm{ydpY}$.

Finally, we have the optimality condition p X p Y = ∂u ∂x ∂u ∂y , which is convenient to write as p X u 2 = p Y u 1 . Differentiating this equation, and lettingu 11 = ∂ 2 u (∂x) 2 , u 12 = ∂ 2 u ∂x∂y , and u 22 = ∂ 2 u (∂y) 2 ,we havep X (u 12 dx+ u 22 dy)= u 1 d p Y + p Y (u 11 dx+ u 12 dy).

For a given dp_Y, we now have three equations in three unknowns: dx, dy, and dM. However, dM only appears in one of the three. Thus, the effect of a price change on x and y can be solved by solving two equations: 0= u 1 dx+ u 2 dy and p X (u 12 dx+ u 22 dy)= u 1 d p Y + p Y (u 11 dx+ u 12 dy) for the two unknowns, dx and dy. This is straightforward and yields

$$\mathrm{dxd}\ pY = -pYu1p \times 2u11 + 2pXpYu12 + pY2u22dydpY = pYu2p \times 2u11 + 2pXpYu12 + pY2u22$$

These equations imply that *x* rises and *y* falls.This is a consequence of the fact that p X 2 u 11 +2 p X p Y u 12 + p Y 2 u 22 <0 , which follows from the already stated second-order condition for a maximum of utility. We immediately see that

$$\mathrm{dy\ dpYdxdpY} = -\mathrm{u1u2} = -\mathrm{pXpY}$$

Thus, the change in (x, y) follows the budget line locally. (This is purely a consequence of holding utility constant.)

To complete the thought while we are embroiled in these derivatives, note that $\mathrm{p} \times \mathrm{u2} = \mathrm{pY\ u\ 1}$ implies that $pXdx + pYdy = 0$

Thus, the amount of money necessary to compensate the consumer for the price increase, keeping utility constant, can be calculated from our third equation:

$$\mathrm{dM} = \mathrm{pXdx} + \mathrm{pY\ dy\ } + \mathrm{ydpY} = \mathrm{ydpY}$$

The amount of income necessary to ensure that the consumer makes no losses from a price increase in Y is the amount that lets him or her buy the bundle that he or she originally purchased; that is, the increase in the amount of money is precisely the amount needed to cover the increased price of y. This shows that locally there is no difference between a substitution effect that keeps utility constant (which is what we explored) and one that provides sufficient income to permit purchasing the previously purchased consumption bundle, at least when small changes in prices are contemplated.

Key Takeaways

- An increase in the price of a good is really a composition of two effects: an increase in the relative price of the good and a decrease in the purchasing power of money. It is useful to examine these two effects separately. The substitution effect considers the change in the relative price, with a sufficient change in income to keep the consumer on the same utility isoquant. The income effect changes only income.
- The substitution effect is the change in consumption resulting from a price change keeping utility constant. The substitution effect always involves a reduction in the good whose price increased.
- The amount of money required to keep the consumer's utility constant from an infinitesimal price increase is precisely the amount required to let him or her buy his or her old bundle at the new prices.

This page titled 12.5: Substitution Effects is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

12.6: Income Effects 12.6: Income EffectsEdit Update titlecancel

Learning Objectives

- How do consumers change their purchases when their income rises or falls?

Wealthy people buy more caviar than poor people. Wealthier people buy more land, medical services, cars, telephones, and computers than poorer people because they have more money to spend on goods and services and, overall, buy more of them. But wealthier people also buy fewer of some goods, too. Rich people buy fewer cigarettes and processed cheese foods. You don't see billionaires waiting in line at McDonald's, and that probably isn't because they have an assistant to wait in line for them. For most goods, at a sufficiently high income, the purchase tends to trail off as income rises.

When an increase in income causes a consumer to buy more of a good, that good is called a normal good for that consumer. When the consumer buys less, the good is called an inferior good, which is an example of sensible jargon that is rare in any discipline. That is, an inferior good is any good whose quantity demanded falls as income rises. At a sufficiently low income, almost all goods are normal goods, while at a sufficiently high income, most goods become inferior. Even a Ferrari is an inferior good against some alternatives, such as Lear jets.

The curve that shows the path of consumption as income changes, holding prices constant, is known as an **Engel curve**.The Engel curve is named for Ernst Engel (1821–1896), a statistician—not for Friedrich Engels, who wrote with Karl Marx. An Engel curve graphs (x(M), y(M)) as M varies, where x(M) is the amount of X chosen with income M, and similarly y(M) is the amount of Y. An example of an Engel curve is illustrated in Figure 12.11.

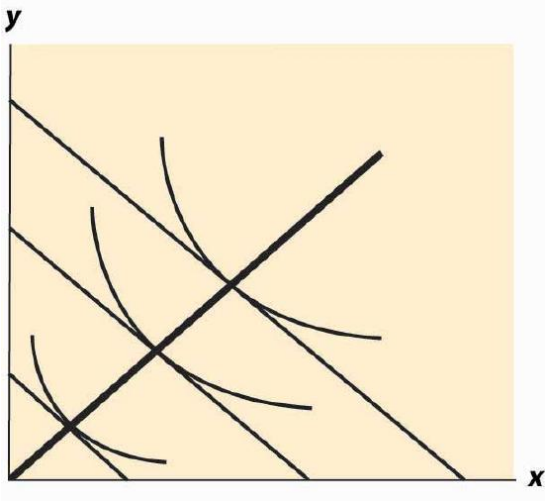


Figure 12.11 Engel curve

Example (Cobb-Douglas): Since the equations $x = aMpX, \quad y = (1-a)MpY$ define the optimal consumption, the Engel curve is a straight line through the origin with slope $(1-a)pXapY$.

An inferior good will see the quantity fall as income rises. Note that, with two goods, at least one is a normal good—they can't both be inferior goods because otherwise, when income rises, less of both would be purchased. An example of an inferior good is illustrated in Figure 12.12. Here, as income rises, the consumption of x rises, reaches a maximum, and then begins to decline. In the declining portion, X is an inferior good.

The definition of the substitution effect now permits us to decompose the effect of a price change into a substitution effect and an income effect. This is illustrated in Figure 12.13.

What is the mathematical form of the income effect? This is actually more straightforward to compute than the substitution effect computed above. As with the substitution effect, we differentiate the conditions $M = pXx + pYy$ and $\mathrm{pXu2} = \mathrm{pY\ u\ 1}$, holding pX and pY constant, to obtain $\mathrm{dM} = \mathrm{pXdx} + \mathrm{pYdy}$ and $pX(u12dx + u22dy) = pY(u11dx + u12dy)$.

Figure 12.12 Backward bending—inferior good

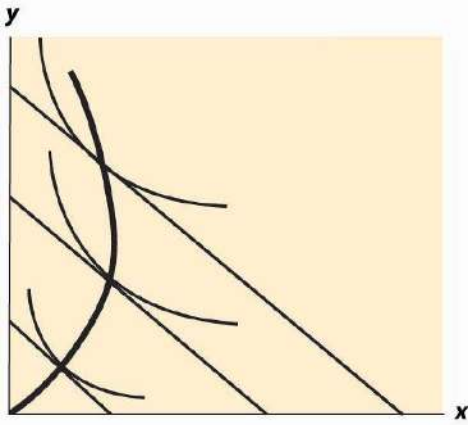
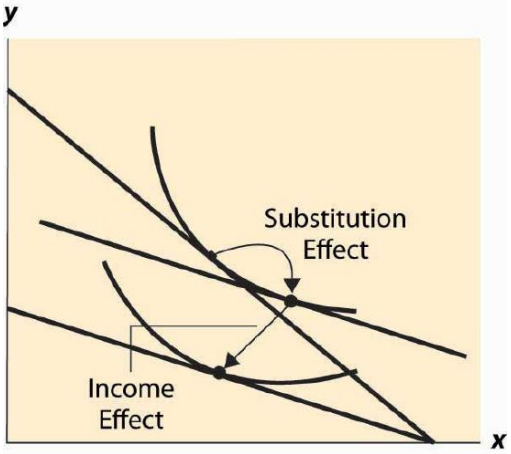


Figure 12.13 Income and substitution effects



The second condition can also be written as $\mathrm{dy\ dx = pY\ u\ 11 - pX\ u\ 12 pX\ u\ 22 - pY\ u\ 12}$

This equation alone defines the slope of the Engel curve without determining how large a change arises from a given change in M. The two conditions together can be solved for the effects of M on X and Y. The Engel curve is given by

$$\mathrm{dxdM = p \times 2u11 - 2pXu12 + p \times 2u22 pXu22 - pY\ u\ 12}$$

and

$$\mathrm{dy\ dM = p\ Y\ 2\ u\ 11 - 2pX\ u\ 12 + p \times 2\ u\ 22\ p\ Y\ u\ 11 - pX\ u\ 12}$$

Note (from the second-order condition) that good Y is inferior if $pYu11 - pXu12 > 0$, or if u 11 u 1 − u 12 u 2 >0, or u 1 u 2 is increasing in x. Since u 1 u 2 is locally constant when M increases, equaling the price ratio, and an increase in y increases u 1 u 2 (thanks to the second-order condition), the only way to keep u 1 u 2 equal to the price ratio is for x to fall. This property characterizes an inferior good—an increase in the quantity of the good increases the marginal rate of substitution of that good for another good.

Key Takeaways

- When an increase in income causes a consumer to buy more of a good, that good is called a normal good for that consumer. When the consumer buys less, the good is called an inferior good. At a sufficiently high income, most goods become inferior.
- The curve that shows the path of consumption as income rises is known as an Engel curve.
- For the Cobb-Douglas utility, Engel curves are straight lines through the origin.
- Not all goods can be inferior.
- The effect of a price increase decomposes into two effects: a decrease in real income and a substitution effect from the change in the price ratio. For normal goods, a price increase decreases quantity. For inferior goods, a price increase decreases quantity only if the substitution effect is larger than the income effect.

EXERCISES

1. Show that, in the case of perfect complements, the Engel curve does not depend on prices.
2. Compute the substitution effect and income effect associated with a multiplicative price increase Δ in p_Y that is, multiplying p_Y by Δ > 1 for the case of the Cobb-Douglas utility $u(x, y) = x \text{ a } y1 - a$

This page titled 12.6: Income Effects is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

12.7: Mathematical Cleanup 12.7: Mathematical CleanupEdit Update titlecancel

Learning Objectives

- Are there important details that haven't been addressed in the presentation of utility maximization?
- What happens when consumers buy none of a good?

Let us revisit the maximization problem considered in this chapter to provide conditions under which local maximization is global. The consumer can spend M on either or both of two goods. This yields a payoff of $h(x) = u(x, M - pXxpY)$. When is this problem well behaved? First, if h is a concave function of x, which implies h ″ (x)≤0, The definition of concavity is such that h is concave if 0 < a < 1 and for all x, y, $h(ax + (1 - a)y) \geq ah(x) + (1 - a)h(y)$. It is reasonably straightforward to show that this implies the second derivative of h is negative; and if h is twice differentiable, the converse is true as well. then any solution to the first-order condition is, in fact, a maximum. To see this, note that h ″ (x)≤0 entails h ′ (x) is decreasing. Moreover, if the point x* satisfies h ′ (x*)=0, then for $x \leq x^*, h'(x) \geq 0; \text{ and for } x \geq x^*, h'(x) \leq 0$, because h ′ (x) gets smaller as x gets larger, and h ′ (x*)=0. Now consider x ≤ x*. Since h ′ (x)≥0, h is increasing as x gets larger. Similarly, for x ≥ x*, h ′ (x)≤0, which means that h gets smaller as x gets larger. Thus, h is concave and h ′ (x*)=0 means that h is maximized at x*.

Thus, a sufficient condition for the first-order condition to characterize the maximum of utility is that h ″ (x)≤0 for all x, p_X, p_Y, and M. Letting z= p X p Y , this is equivalent to u 11 −2z u 12 + z 2 u 22 ≤0 for all z > 0.

In turn, we can see that this requires (i) $u_{11} \leq 0$ (z = 0), (ii) $u_{22} \leq 0$ (z→∞), and (iii) u 11 u 22 − u 12 ≥0 (z= u 11 u 22). In addition, since

$$-(u11 + 2zu12 + z2u22) = (-u11 - z - u22)2 + 2z(u11u22 - u12)$$

(i), (ii), and (iii) are sufficient for u11 +2zu 12 + z2u 22 ≤0.

Therefore, if (i) $u_{11} \leq 0$, (ii) $u_{22} \leq 0$, and (iii) u 11 u 22 − u 12 ≥0, a solution to the first-order conditions characterizes utility maximization for the consumer.

When will a consumer specialize and consume zero of a good? A necessary condition for the choice of x to be zero is that the consumer doesn't benefit from consuming a very small x ; that is, h ′ (0)≤0. This means that

$$h'(0) = u1(0, MpY) - u2(0, MpY)pXpY \leq 0$$

or

$$\text{u } 1(0, \text{MpY}) \text{ u } 2(0, \text{MpY}) \leq \text{pXpY}$$

Moreover, if the concavity of h is met, as assumed above, then this condition is sufficient to guarantee that the solution is zero. To see this, note that concavity of h implies h ′ is decreasing. Combined with h ′ (0)≤0, this entails that h is maximized at 0. An important class of examples of this behavior is quasilinear utility. Quasilinear utility comes in the form u(x, y) = y + v(x), where v is a concave function (v ″ (x)≤0 for all x). That is, **quasilinear utility** is utility that is additively separable.

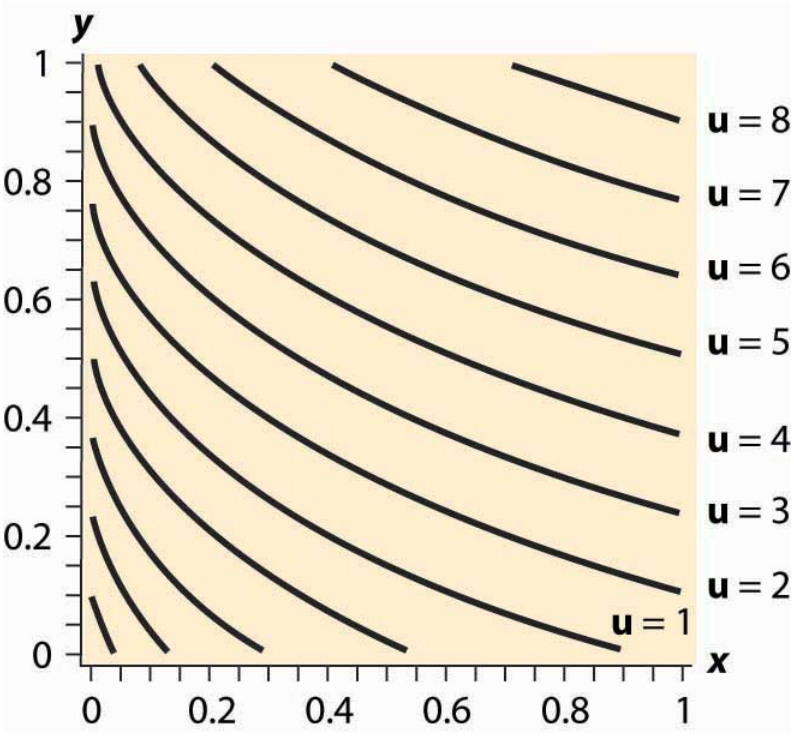


Figure 12.14 Quasilinear isoquants

The procedure for dealing with corners is generally this. First, check concavity of the h function. If h is concave, we have a procedure to solve the problem; when h is not concave, an alternative strategy must be devised. There are known strategies for some cases that are beyond the scope of this text. Given h concave, the next step is to check the endpoints and verify that h ′ (0)>0 (for otherwise x = 0 maximizes the consumer's utility) and h ′ (M p X)<0 (for otherwise y = 0 maximizes the consumer's utility). Finally, at this point we seek the interior solution h ′ (x)=0. With this procedure, we can ensure that we find the actual maximum for the consumer rather than a solution to the first-order conditions that don't maximize the consumer's utility.

Key Takeaways

- Conditions are available that ensure that the first-order conditions produce a utility maximum.
- With convex preferences, zero consumption of one good arises when utility is decreasing in the consumption of one good, spending the rest of income on the other good.

EXERCISE

1. Demonstrate that the quasilinear consumer will consume zero X if and only if $v'(0) \leq pxpy$, and that the consumer instead consumes zero Y if $v'(MpX) \geq pxpy$. The quasilinear utility isoquants, for $v(x) = (x + 0.03)0.3$, are illustrated in Figure 12.14. Note that, even though the isoquants curve, they are nonetheless parallel to each other.

This page titled 12.7: Mathematical Cleanup is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

13: Applied Consumer Theory

13.1: Labor Supply
13.2: Urban Real Estate Prices
13.3: Dynamic Choice
13.4: Risk Aversion
13.5: Search

This page titled 13: Applied Consumer Theory is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

13.1: Labor Supply 13.1: Labor SupplyEdit Update titlecancel

Learning Objectives

- If we want people to work more, should we pay them more or will that cause them to work less?

Consider a taxi driver who owns a car, a convenience store owner, or anyone else who can set his or her own hours. Working has two effects on this consumer: more goods consumption but less leisure consumption. To model this, we let x be the goods consumption, L the amount of nonwork time or leisure, and working time T – L, where T is the amount of time available for activities of all kinds. The variable L includes a lot of activities that aren't necessarily fun—like trips to the dentist, haircuts, and sleeping—but for which the consumer isn't paid and which represent choices. One could argue that sleeping isn't really a choice, in the sense that one can't choose zero sleep; but this can be handled by adjusting T to represent "time available for chosen behavior" so that T – L is work time and L is chosen nonwork activities. We set L to be leisure rather than labor supply because it is leisure that is the good thing, whereas most of us view working as something that we are willing to do provided we're paid for it.

Labor supply is different from other consumption because the wage enters the budget constraint twice: first as the price of leisure, and second as income from working. One way of expressing this is to write the consumer's budget constraint as $px + wL = M + wT$.

Here, M represents nonwork income, such as gifts, government transfers, and interest income. We drop the subscript on the price of X and use w as the wage. Finally, we use a capital L for leisure because a lowercase L looks like the number one. The somewhat Dickensian idea is that the consumer's maximal budget entails working the total available hours T, and any nonworked hours are purchased at the wage rate w. Alternatively, one could express the budget constraint so as to reflect that expenditures on goods px equal the total money, which is the sum of nonwork income M and work income $w(T - L),$ or $px = M + w(T - L)$.

These two formulations of the budget constraint are mathematically equivalent.

The strategy for solving the problem is also equivalent to the standard formulation, although there is some expositional clarity used by employing the budget constraint to eliminate x. That is, we write the utility $u(x, L)$ as $h(L) = u(M + w(T - L)pL)$.

As before, we obtain the first-order condition $0 = h'(L^*) = -u1(wp) + u2$, where the partial derivatives u_1 and u_2 are evaluated at $(M + w(T - L^*)p, L^*)$. Note that the first-order condition is the same as the standard two-good theory developed already. This is because the effect, so far, is merely to require two components to income: M and wT, both of which are constant. It is only when we evaluate the effect of a wage increase that we see a difference.

To evaluate the effect of a wage increase, differentiate the first-order condition to obtain

$$0 = (u11(wp)2 - 2u12(wp) + u22)dLdw - u1p - (wp)u11T - Lp + u12T - Lp$$

Since $u\ 11(\ w\ p)2 - 2\ u\ 12(wp) + u22 < 0$

by the standard second-order condition,

dL dw >0 if and only if u 1 p +(w p) u 11 T−L p − u 12 T−L p <0;

that is, these expressions are equivalent to one another. Simplifying the latter, we obtain −(w p) u 11 T−L p + u 12 T−L p u 1 p >1, or (T−L) −(w p) u 11 + u 12 u 1 >1, or ∂ ∂L Log(u 1)> 1 T−L =− ∂ ∂L Log(T−L), or ∂ ∂L Log(u 1)+ ∂ ∂L Log(T−L)>0, or ∂ ∂L Log(u 1 (T−L))>0.

Since the logarithm is increasing, this is equivalent to u 1 (T−L) being an increasing function of L. That is, L rises with an increase in wages and a decrease in hours worked if the marginal utility of goods times the hours worked is an increasing function of L, holding constant everything else, but evaluated at the optimal values. The value u_1 is the marginal value of an additional good, while the value T – L represents the hours worked. Thus, in particular, if goods and leisure are substitutes, so that an increase in L decreases the marginal value of goods, then an increase in wages must decrease leisure, and labor supply increases in wages. The case where the goods are complements holds a hope for a decreasing labor supply, so we consider first the extreme case of complements.

Example (Perfect complements): $u(x, L) = \min\{x, L\}$

In this case, the consumer will make consumption and leisure equal to maximize the utility, so $\mathrm{M} + \mathrm{w}\,(\mathrm{T} - \mathrm{L}^*)\,\mathrm{p} = \mathrm{L}^*$ or $\mathrm{L}^* = \mathrm{M} + \mathrm{wTp1} + \mathrm{wp} = \mathrm{M} + \mathrm{wTp} + \mathrm{w}$.

Thus, L is increasing in the wages if pT > M; that is, if M is sufficiently small so that one can't buy all of one's needs and not work at all. (This is the only reasonable case for this utility function.) With strong complements between goods and leisure, an increase in wages induces fewer hours worked.

Example (Cobb-Douglas): $h(L) = (M + w(T - L)p)aL1 - a.$

The first-order condition gives $0 = h'(L) = -a(M + w(T - L)p)a - 1\lfloor 1 - awp + (1 - a)(M + w(T - L)p)aL - a$

or $\text{al}\ wp = (1 - a)M + w(T - L)p,$

$\text{wpL} = (1 - \text{a})\text{M} + \text{wTp}$

or $\text{L} = (1 - \text{a})(\text{Mw} + \text{T})$

If M is high enough, the consumer doesn't work but takes L = T; otherwise, the equation gives the leisure, and labor supply is given by $\text{T}-\text{L}=\text{Max}\{\ 0,\alpha\text{T}-(1-\alpha)(\ \text{M w}\ )\ \}.$

Labor supply increases with the wage, no matter how high the wage goes.

The wage affects not just the price of leisure but also the income level. This makes it possible for the income effect of a wage increase to dominate the substitution effect. Moreover, we saw that this is more likely when the consumption of goods takes time; that is, the goods and leisure are complements.

Figure 13.1 Hours per week

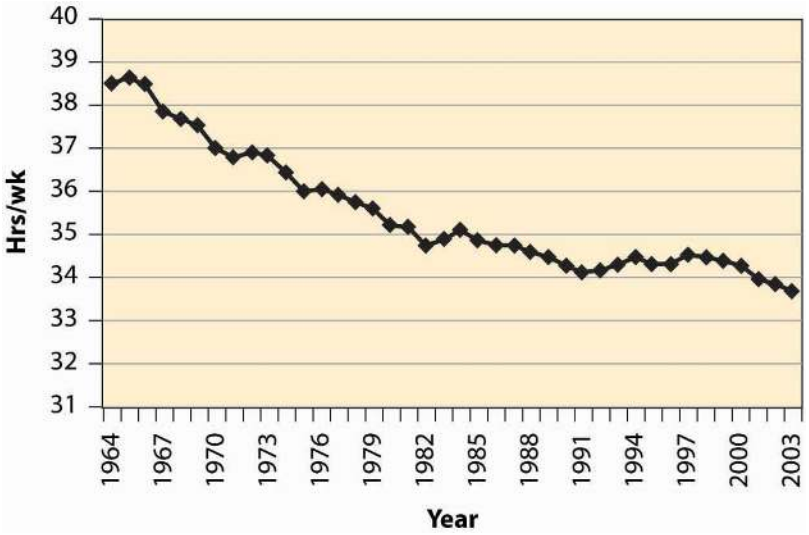


Figure 13.1 illustrates.

A number of physicists have changed careers to become researchers in finance or financial economics. Research in finance pays substantially better than research in physics, and yet requires many of the same mathematical skills like stochastic calculus. Physicists who see their former colleagues driving Porsches and buying summerhouses are understandably annoyed that research in finance—which is intellectually no more difficult or challenging than physics—pays so much better. Indeed, some physicists are saying that other fields—such as finance, economics, and law—"shouldn't" pay more than physics.

The difference in income between physics' researchers and finance researchers is an example of a **compensating differential**. A compensating differential is income or costs that equalize different choices. There are individuals who could become either physicists or finance researchers. At equal income, too many choose physics and too few choose finance, in the sense that there is a surplus of physicists and a shortage of finance researchers. Finance salaries must exceed physics' salaries in order to induce some of the researchers who are capable of doing either one to switch to finance, which compensates those individuals for doing the less desirable task.

Jobs that are dangerous or unpleasant must pay more than jobs requiring similar skills but without the bad attributes. Thus, oil-field workers in Alaska's North Slope, well above the Arctic Circle, earn a premium over workers in similar jobs in Houston, Texas. The premium—or differential pay—must be such that the marginal worker is indifferent between the two choices: The extra pay compensates the worker for the adverse working conditions. This is why it is known in economics' jargon by the phrase of a compensating differential.

The high salaries earned by professional basketball players are not compensating differentials. These salaries are not created because of a need to induce tall people to choose basketball over alternative jobs like painting ceilings, but instead are payments that reflect the rarity of the skills and abilities involved. Compensating differentials are determined by alternatives, not by direct scarcity. Professional basketball players are well paid for the same reason that Picasso's paintings are expensive: There aren't very many of them relative to demand.

A compensating differential is a feature of other choices as well as career choices. For example, many people would like to live in California for its weather and scenic beauty. Given the desirability of California over, for example, Lincoln, Nebraska, or Rochester, New York, there must be a compensating differential for living in Rochester; and two significant ones are air quality and housing prices. Air quality worsens as populations rise, thus tending to create a compensating differential. In addition, the increase in housing prices also tends to compensate—housing is inexpensive in Rochester, at least compared with California.There are other compensations, besides housing, for living in Rochester—cross-country skiing and proximity to mountains and lakes, for example. Generally, employment is only a temporary factor that might compensate, because employment tends to be mobile, too, and move to the location that the workers prefer, when possible. It is not possible on Alaska's North Slope.

Housing prices also compensate for location within a city. For most people, it is more convenient—both in commuting time and for services—to be located near the central business district than in the outlying suburbs. The main compensating differentials are school quality, crime rates, and housing prices. We illustrate the ideas with a simple model of a city in the next section.

Key Takeaways

- Leisure—time spent not working—is a good like other goods, and the utility cost of working is less leisure.
- Labor supply is different from other goods because the wage enters the budget constraint twice—first as the price of leisure, and second as income from working.
- If goods and leisure are substitutes, so that an increase in L decreases the marginal value of goods, then an increase in wages must decrease leisure, and labor supply increases in wages.
- With strong complements between goods and leisure, an increase in wages induces fewer hours worked.
- Complementarity between goods and leisure is reasonable because it takes time to consume goods.
- For most developed nations, increases in wages are associated with fewer hours worked.
- A compensating differential is income or costs that equalize different choices.
- Jobs that are dangerous or unpleasant must pay more than jobs requiring similar skills but without the bad attributes.
- The premium—or differential pay—must be such that the marginal worker is indifferent between the two choices: The extra pay compensates the worker for the adverse working conditions.
- City choice is also subject to compensating differentials, and significant differentials include air quality, crime rates, tax rates, and housing prices.

EXERCISES

1. A thought question: Does a bequest motive—the desire to give money to others—change the likelihood that goods and leisure are complements?
2. Show that an increase in the wage increases the consumption of goods; that is, x increases when the wage increases.

This page titled 13.1: Labor Supply is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

13.2: Urban Real Estate Prices 13.2: Urban Real Estate PricesEdit Update titlecancel

Learning Objectives

- How are the prices of suburban ranch houses, downtown apartments, and rural ranches determined?

An important point to understand is that the good, in limited supply in cities, is not a physical structure like a house, but the land on which the house sits. The cost of building a house in Los Angeles is quite similar to the cost of building a house in Rochester, New York. The big difference is the price of land. A $1 million house in Los Angeles might be a $400,000 house sitting on a $600,000 parcel of land. The same house in Rochester might be $500,000—a $400,000 house on a $100,000 parcel of land.

Usually, land is what fluctuates in value rather than the price of the house that sits on the land. When a newspaper reports that house prices rose, in fact what rose were land prices, for the price of housing has changed only at a slow pace, reflecting increased wages of house builders and changes in the price of lumber and other inputs. These do change, but historically the changes have been small compared to the price of land.

We can construct a simple model of a city to illustrate the determination of land prices. Suppose the city is constructed on a flat plane. People work at the origin (0, 0). This simplifying assumption is intended to capture the fact that a relatively small, central portion of most cities involves business, with a large area given over to housing. The assumption is extreme, but not unreasonable as a description of some cities.

Suppose commuting times are proportional to distance from the origin. Let c(t) be the cost to the person of a commute of time t, and let the time taken be t = λr, where r is the distance. The function c should reflect both the transportation costs and the value of time lost. The parameter λ accounts for the inverse of the speed in commuting, with a higher λ indicating slower commuting. In addition, we assume that people occupy a constant amount of land. This assumption is clearly wrong empirically, and we will consider making house size a choice variable later.

A person choosing a house priced at p (r), at distance r, thus pays $c(\lambda r) + p(r)$ for the combination of housing and transportation. People will choose the lowest cost alternative. If people have identical preferences about housing and commuting, then house prices p will depend on distance and will be determined by $c(\lambda r) + p(r)$ equal to a constant, so that people are indifferent to the distance from the city's center—decreased commute time is exactly compensated by increased house prices.

The remaining piece of the model is to figure out the constant. To do this, we need to figure out the area of the city. If the total population is N, and people occupy an area of one per person, then the city size r_{max} satisfies $\mathrm{N} = \mathrm{\Pi r}\max 2$, and thus $r\max = N\Pi$.

At the edge of the city, the value of land is given by some other use, like agriculture. From the perspective of the determinant of the city's prices, this value is approximately constant. As the city takes more land, the change in agricultural land is a very small portion of the total land used for agriculture. Let the value of agricultural land be v per housing unit size. Then the price of housing is $p\left(r_{\max}\right) = v$, because this is the value of land at the edge of the city. This lets us compute the price of all housing in the city:

$$c(\lambda r) + p(r) = c(\lambda r\max) + p(r\max) = c(\lambda r\max) + v = c(\lambda Nn) + v$$

or

$$p(r) = c(\lambda Nn) + v - c(\lambda r)$$

This equation produces housing prices like those illustrated in Figure 13.2, where the peak is the city's center. The height of the figure indicates the price of housing.

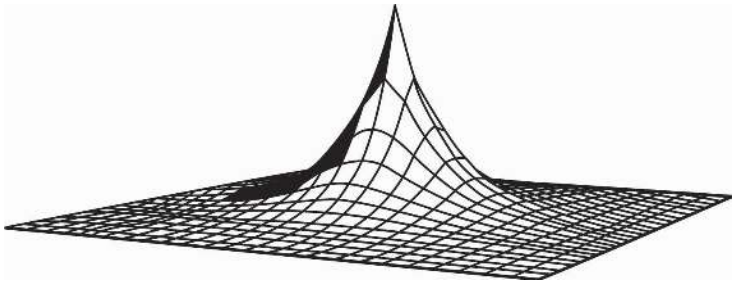


Figure 13.2 House price gradient

It is straightforward to verify that house prices increase in the population N and the commuting time parameter λ, as one would expect. To quantify the predictions, we consider a city with a population of 1,000,000; a population density of 10,000 per square mile; and an agricultural use value of $6 million per square mile. To translate these assumptions into the model's structure, first note that a population density of 10,000 per square mile creates a fictitious "unit of measure" of about 52.8 feet, which we'll call a purlong, so that there is one person per square purlong (2,788 square feet). Then the agricultural value of a property is v = $600 per square purlong. Note that this density requires a city of radius r_{max} equal to 564 purlongs, which is 5.64 miles.

The only remaining structure left to identify in the model is the commuting cost c. To simplify the calculations, let c be linear. Suppose that the daily cost of commuting is $2 per mile (roundtrip), so that the present value of daily commuting costs in perpetuity is about $10,000 per mile.This amount is based upon 250 working days per year, for an annual cost of about $500 per mile, yielding a present value at 5% interest of $10,000. See Section 11.1. With a time value of $25 per hour and an average speed of 40 mph (1.5 minutes per mile), the time cost is 62.5 cents per minute. Automobile costs (such as gasoline, car depreciation, and insurance) are about 35–40 cents per mile. Thus the total is around $1 per mile, which doubles with roundtrips. This translates into a cost of commuting of $100 per purlong. Thus, we obtain

$$\mathrm{p(r) = c(\lambda Nn) + v - c(\lambda r) = \$100(Nn - r) + \$600 = \$57,000 - \$100r}$$

Thus, the same 2,788-square-foot property at the city's edge sells for $600 versus $57,000, less than six miles away at the city's center. With reasonable parameters, this model readily creates dramatic differences in land prices, based purely on commuting time.

As constructed, a quadrupling of population approximately doubles the price of land in the central city. This probably understates the change, since a doubling of the population would likely increase road congestion, increasing λ and further increasing the price of central city real estate.

As presented, the model contains three major unrealistic assumptions. First, everyone lives on an identically sized piece of land. In fact, however, the amount of land used tends to fall as prices rise. At $53 per square foot, most of us buy a lot less land than at 20 cents per square foot. As a practical matter, the reduction of land per capita is accomplished both through smaller housing units and through taller buildings, which produce more housing floor space per acre of land. Second, people have distinct preferences and the disutility of commuting, as well as the value of increased space, varies with the individual. Third, congestion levels are generally endogenous—the more people who live between two points, the greater the traffic density and, consequently, the higher the level of λ. Problems arise with the first two assumptions because of the simplistic nature of consumer preferences embedded in the model, while the third assumption presents an equilibrium issue requiring consideration of transportation choices.

This model can readily be extended to incorporate different types of people, different housing sizes, and endogenous congestion. To illustrate such generalizations, consider making the housing size endogenous. Suppose preferences are represented by the utility function $\mathrm{u = Ha - \lambda r - p(r)H}$, where H is the house size that the person chooses, and r is the distance that he or she chooses. This adaptation of the model reflects two issues. First, the transport cost has been set to be linear in distance for simplicity. Second, the marginal value of housing decreases in the house size, but the value of housing doesn't depend on distance from the center. For these preferences to make sense, $a < 1$ (otherwise either zero or an infinite house size emerges). A person with these preferences would optimally choose a house size of $\mathrm{H = (ap(r))11 - a}$, resulting in utility

$u^* = (a\ \mathrm{a}\ 1 - a - a11 - a)p(r) - a1 - a - \lambda r$. Utility at every location is constant, so $(aa1 - a - a11 - au^* + \lambda r)\,1 - aa = p(r)$.

A valuable attribute of the form of the equation for p is that the general form depends on the equilibrium values only through the single number u*. This functional form produces the same qualitative shapes as shown in Figure 13.2. Using the form, we can solve for the housing size H:

H(r)= (α p(r)) 1 1−α = α 1 1−α (u*+λr α α 1−α − α 1 1−α) 1 α = (u*+λr α −1 −1) 1 α = (α 1−α (u*+λr)) 1 α .

The space in the interval $[r, r + \Delta]$ is $n\,(2r\Delta + \Delta^2)$. In this interval, there are approximately $n(2r\Delta + \Delta 2)H(r) = n(2r\Delta + \Delta 2)\,(1 - aa\,(u^* + \lambda r))\,1$ α people. Thus, the number of people within r_{max} of the city's center is $\int 0\ \mathrm{r}\ \max 2\pi r\,(1 - aa\,(u^* + \lambda r))\,1\ \mathrm{a}\ \mathrm{dr} = N$

This equation, when combined with the value of land on the periphery $\mathrm{v = p(r\,max) = (aa1 - a - a11 - au^* + \lambda r\,max)\,1 - aa}$, jointly determines r_{max} and u*.

When different people have different preferences, the people with the highest disutility of commuting will tend to live closer to the city's center. These tend to be people with the highest wages, since one of the costs of commuting is time that could have been spent working.

Key Takeaways

- An important point to understand is that, in cities, houses are not in limited supply; but it is the land on which the houses sit that is.
- The circular city model involves people who work at a single point but live dispersed around that point. It is both the size of the city and the housing prices that are determined by consumers who are indifferent to commuting costs—lower housing prices at a greater distance balance the increased commuting costs.
- Substituting plausible parameters into the circular city model produces dramatic house price differentials, explaining much of the price differences within cities.
- A quadrupling of population approximately doubles the price of land in the central city. This likely understates the actual change since an increase in population slows traffic.

EXERCISE

1. For the case of $a = 1/2$, solve for the equilibrium values of u* and r_{max}.

This page titled 13.2: Urban Real Estate Prices is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

13.3: Dynamic Choice 13.3: Dynamic ChoiceEdit Update titlecancel

Learning Objectives

- How much should you save, and how do interest rate changes affect spending?

The consumption of goods doesn't take place in a single instance, but over time. How does time enter into choice? We're going to simplify the problem a bit and focus only on consumption, setting aside working for the time being. Let x_1 be consumption in the first period and x_2 in the second period. Suppose the value of consumption is the same in each period, so that $u(x_1, x_2) = v(x_1) + \delta v(x_2)$, where δ is called the rate of "pure" time preference. The consumer is expected to have income M_1 in the first period and M_2 in the second. There is a market for lending and borrowing, which we assume has a common interest rate r.

The consumer's budget constraint, then, can be written $(1 + r)(M_1 - x_1) = x_2 - M_2$

This equation says that the net savings in Period 1, plus the interest on the net savings in Period 1, equals the net expenditures in Period 2. This is because whatever is saved in Period 1 earns interest and can then be spent in Period 2; alternatively, whatever is borrowed in Period 1 must be paid back with interest in Period 2. Rewriting the constraint: $(1+r)x_1 + x_2 = (1+r)M_1 + M_2$

This equation is known as the **intertemporal budget constraint**—that is, the budget constraint that allows for borrowing or lending. It has two immediate consequences. First, 1 + r is the price of Period 2 consumption in terms of Period 1 consumption. Thus, the interest rate gives the relative prices. Second, the relevant income is the **permanent income**, which is the present value of the income stream. Clearly a change in income that leaves the present value of income the same should have no effect on the choice of consumption.

Once again, as with the labor supply, a change in the interest rate affects not just the price of consumption but also the budget for consumption. In other words, an increase in the interest rate represents an increase in budget for net savers but a decrease in budget for net borrowers.

As always, we rewrite the optimization problem to eliminate one of the variables, to obtain $u = v(x1) + \delta v((1+r)(M1 - x1) + M2)$

Thus, the first-order conditions yield $0 = v'(x1) - (1+r)\delta v'(x2)$

This condition says that the marginal value of consumption in Period 1, v ′ (x 1), equals the marginal value of consumption in Period 2, δ v ′ (x 2), times the interest factor. That is, the marginal present values are equated. Note that the consumer's private time preference, δ, need not be related to the interest rate. If the consumer values Period 1 consumption more than does the market, so that $\delta(1+r) < 1$, then $v'(x1) < v'(x2)$; that is, the consumer consumes more in Period 1 than in Period 2.As usual, we are assuming that utility is concave, which in this instance means that the second derivative of v is negative, and thus the derivative of v is decreasing. In addition, to ensure an interior solution, it is useful to require what are called the Inada conditions: $v'(0) = \infty$ and $v'(\infty) = 0$ Similarly, if the consumer's discount of future consumption is exactly equal to the market discount, $\delta(1+r) = 1$, the consumer will consume the same amount in both periods. Finally, if the consumer values Period 1 consumption less than the market, $\delta(1+r) > 1$, the consumer will consume more in Period 2. In this case, the consumer is more patient than the market.

Figure 13.3 Borrowing and lending

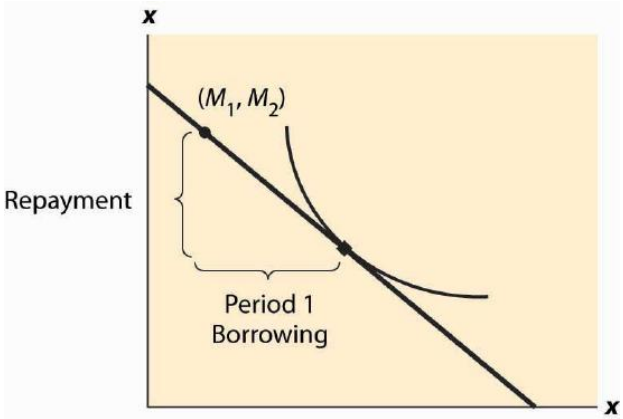


Figure 13.3. In this figure, the consumer's income mostly comes in the second period. As a consequence, the consumer borrows in the first period and repays in the second period.

The effect of an interest rate increase is to pivot the budget constraint around the point (M_1, M_2). Note that this point is always feasible—that is, it is feasible to consume one's own endowment. The effect of an increase in the interest rate is going to depend on whether the consumer is a borrower or a lender. As Figure 13.4 illustrates, the net borrower borrows less in the first period—the price of first-period consumption has risen and the borrower's wealth has fallen. It is not clear whether the borrower consumes less in the second period because the price of second-period consumption has fallen, even though wealth has fallen, too—two conflicting effects.

An increase in interest rates is a benefit to a net lender. The lender has more income, and the price of Period 2 consumption has fallen. Thus the lender must consume more in the second period, but only consumes more in the first period (lends less) if the income effect outweighs the substitution effect. This is illustrated in Figure 13.5.

Figure 13.4 Interest rate change

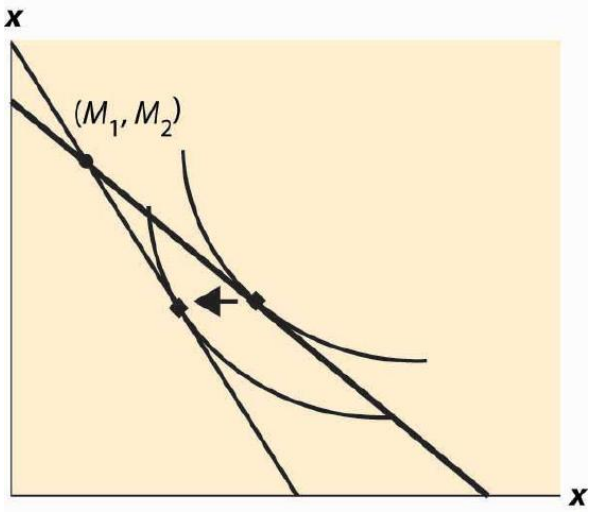
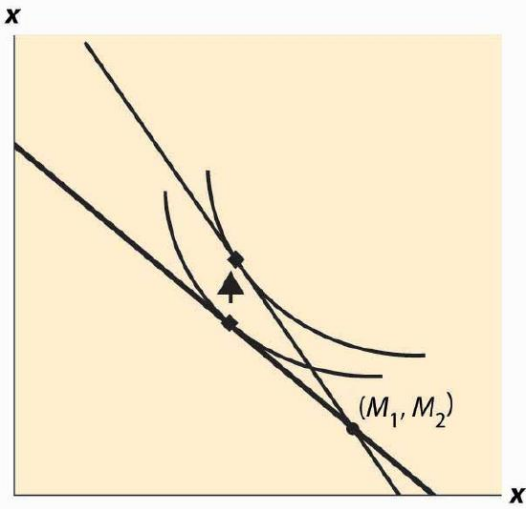


Figure 13.5 Interest rate increase on lenders



The government, from time to time, will rebate a portion of taxes to "stimulate" the economy. An important aspect of the outcome of such a tax rebate is the effect to which consumers will spend the rebate, versus save the rebate, because the stimulative effects of spending are thought to be larger than the stimulative effects of saving.This belief shouldn't be accepted as necessarily true; it was based on a model that has since been widely rejected by the majority of economists. The general idea is that spending creates demand for goods, thus encouraging business investment in production. However, saving encourages investment by producing spendable funds, so it isn't at all obvious whether spending or saving has a larger effect. The theory suggests how people will react to a "one-time" or transitory tax rebate, compared to a permanent lowering of taxes. In particular, the budget constraint for the consumer spreads lifetime income over the lifetime. Thus, for an average consumer who might spend a present value of $750,000

over a lifetime, a $1,000 rebate is small potatoes. On the other hand, a $1,000 per year reduction is worth $20,000 or so over the lifetime, which should have 20 times the effect of the transitory change on the current expenditure.

Tax rebates are not the only way that we receive one-time payments. Money can be found, or lost, and we can have unexpected costs or windfall gifts. From an intertemporal budget constraint perspective, these transitory effects have little significance; and thus the theory suggests that people shouldn't spend much of a windfall gain in the current year, or cut back significantly when they have a moderately sized, unexpected cost.

As a practical matter, most individuals can't borrow at the same rate at which they lend. Many students borrow on credit cards at very high interest rates and obtain a fraction of that in interest on savings. That is to say, borrowers and lenders face different interest rates. This situation is readily identified in Figure 13.6. The cost of a first-period loan is a relatively high loss of x_2, and similarly the value of first-period savings is a much more modest increase in second-period consumption. Such effects tend to favor "neither a borrower nor a lender be," as Shakespeare recommends, although it is still possible for the consumer to optimally borrow in the first period (e.g., if $M_1 = 0$) or in the second period (if M_2 is small relative to M_1).

Figure 13.6 Different rates for borrowing and lending

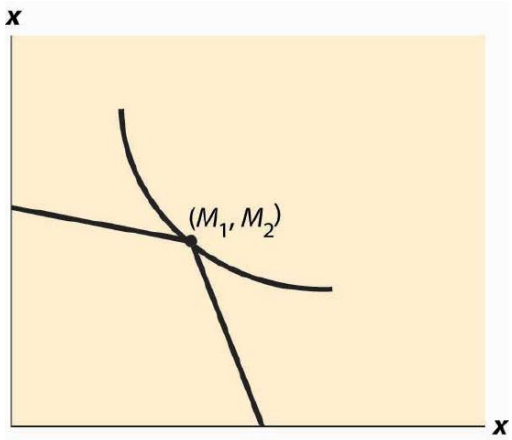


Figure 13.7.

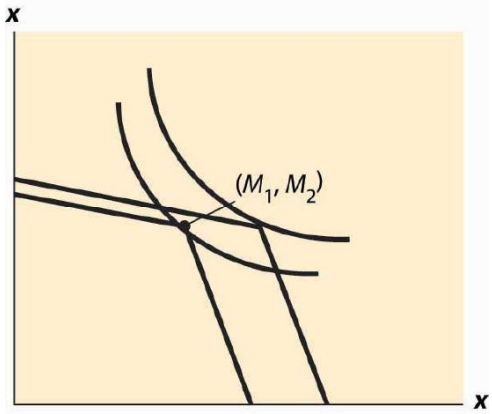


Figure 13.7 The effect of a transitory income increase

Key Takeaways

- The intertemporal budget constraint takes into account the fact that savings obtain interest, producing additional money. The price of early consumption is one plus the interest rate.
- The relevant income is "permanent income" rather than "current income." A change in income that leaves the present value of income the same should have no effect on the choice of consumption.
- A change in the interest rate affects not just the price of consumption but also the budget for consumption. An increase in the interest rate represents an increase in budget for savers but a decrease in budget for borrowers.
- If the consumer values early consumption more than the market, the consumer consumes more early rather than later, and conversely.

- Whether the consumer is a lender or borrower depends not just on the preference for earlier versus later consumption but also on the timing of their incomes.
- The effect of an interest rate increase is to pivot the budget constraint around the income point. The effect of an increase in the interest rate on consumption is going to depend on whether the consumer is a borrower or a lender.
- An increase in interest rates is a benefit to a net lender. The lender must continue to lend in the present and will consume more in the future.
- People should react less to a “one-time” or transitory tax rebate than to a permanent lowering of taxes.
- Most individuals can’t borrow at the same rate at which they lend. Interest rate differentials favor spending all of one’s income.
- Differences in borrowing and lending interest rates cause transitory changes in income to have larger effects than the intertemporal budget constraint would suggest.

This page titled 13.3: Dynamic Choice is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

13.4: Risk Aversion 13.4: Risk AversionEdit Update titlecancel

Learning Objectives

- How should you evaluate gambles?
- How is risk priced?

There are many risks in life, even if one doesn't add to these risks by intentionally buying lottery tickets. Gasoline prices go up and down, the demand for people trained in your major fluctuates, and house prices change. How do people value gambles? The starting point for the investigation is the von Neumann-MorgensternJohn von Neumann (1903–1957) and Oskar Morgenstern (1902–1977) are the authors of Theory of Games and Economic Behavior (Princeton, NJ: Princeton University Press, 1944). utility function. The idea of a **von Neumann-Morgenstern utility function** for a given person is that, for each possible outcome x, there is a value v(x) assigned by the person, and the average value of v is the value the person assigns to the risky outcome. In other words, the von Neumann-Morgenstern utility function is constructed in such a way that a consumer values gambles as if they were the expected utility

This is a "state-of-the-world" approach, in the sense that each of the outcomes is associated with a state of the world, and the person maximizes the expected value of the various possible states of the world. Value here doesn't mean a money value, but a psychic value or utility.

To illustrate the assumption, consider equal probabilities of winning $100 and winning $200. The expected outcome of this gamble is $150—the average of $100 and $200. However, the actual value of the outcome could be anything between the value of $100 and the value of $200. The von Neumann-Morgenstern utility is $1/2v(\$100) + 1/2v(\$200)$

The von Neumann-Morgenstern formulation has certain advantages, including the logic that what matters is the average value of the outcome. On the other hand, in many tests, people behave in ways not consistent with the theory.For example, people tend to react more strongly to very unlikely events than is consistent with the theory. Nevertheless, the von Neumann approach is the prevailing model of behavior under risk.

To introduce the theory, we will consider only money outcomes, and mostly the case of two money outcomes. The person has a von Neumann-Morgenstern utility function v of these outcomes. If the possible outcomes are x_1, x_2, … , x_n and these occur with probability π_1, π_2, … , π_n respectively, the consumer's utility is

$$u = \pi 1v(x1) + \pi 2v(x2) + ... + \pi nv(xn) = \sum i = 1n\pi iv(xi).$$

This is the meaning of "having a von Neumann-Morgenstern utility function"—that utility can be written in this weighted sum form.

The first insight that flows from this definition is that an individual dislikes risk if v is concave. To see this, note that the definition of concavity posits that v is concave if, for all π in [0, 1] and all values x_1 and x_2, $v(n \times 1 + (1-n) \times 2) \geq \pi v(\times 1) + (1-n)v(\times 2)$

For smoothly differentiable functions, concavity is equivalent to a second derivative that is not positive. Using induction, the definition of concavity can be generalized to show $v(\pi 1x1 + \pi 2x2 + ... + \pi nxn) \geq \pi 1v(x1) + \pi 2v(x2) + ... + \pi nv(xn)$..

Figure 13.8 Expected utility and certainty equivalents

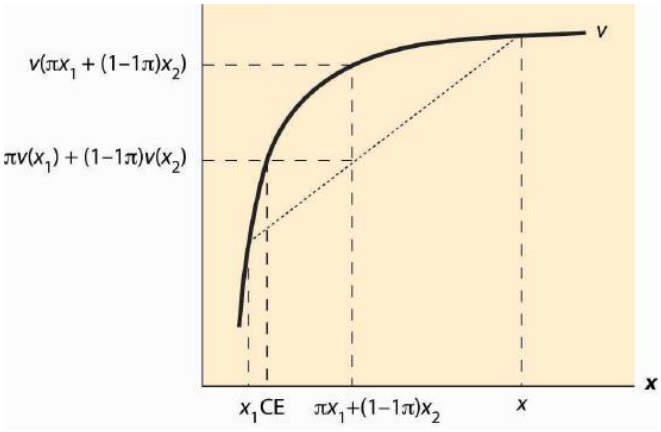


Figure 13.8. There are two possible outcomes: x_1 and x_2. The value x_1 occurs with probability π, and x_2 with probability 1 – π. This means that the average or expected outcome is $\pi x1 + (1–\pi)x2$. The value $v(\pi x1 + (1–\pi)x2)$ is the value at the expected outcome $\pi x1 + (1–\pi)x2, while \pi v(x1) + (1–\pi)v(x2)$ is the average of the value of the outcome. As is plainly visible in the figure, concavity makes the average outcome preferable to the random outcome. People with concave von Neumann-Morgenstern utility functions are known as **risk-averse** people—they prefer the expected value of a gamble to the gamble itself.

A useful concept is the certainty equivalent of a gamble. The **certainty equivalent** is an amount of money that provides equal utility to the random payoff of the gamble. The certainty equivalent is labeled CE in the figure. Note that CE is less than the expected outcome, if the person is risk averse. This is because risk-averse individuals prefer the expected outcome to the risky outcome.

The **risk premium** is defined to be the difference between the expected payoff (this is expressed as πx1 + (1 – π)x2 in the figure) and the certainty equivalent. This is the cost of risk—it is the amount of money an individual would be willing to pay to avoid risk. This means as well that the risk premium is the value of insurance. How does the risk premium of a given gamble change when the base wealth is increased? It can be shown that the risk premium falls as wealth increases for any gamble, if and only if $-v^n(x)v'(x)$ is decreasing.

The measure $\rho(x) = -v''(x)v'(x)\backslash](x)$ is known as the **Arrow-Pratt measure of risk aversion**The measure was named after its discoverers, Nobel laureate Kenneth Arrow and John Pratt., and also as the measure of absolute risk aversion. It is a measure of risk aversion computed as the negative of the ratio of the second derivative of utility divided by the first derivative of utility. To get an idea about why this measure matters, consider a quadratic approximation to v. Let μ be the expected value, and let δ^2 be the expected value of \((x-\mu)^{2} \nonumber \]

). Then we can approximate v(CE) two different ways.

$$\mathrm{v}(\mu) + \mathrm{v}'(\mu)(\mathrm{CE} - \mu) \approx \mathrm{v}(\mathrm{CE}) = \mathrm{E}\{\mathrm{v}(\mathrm{x})\} \approx \mathrm{E}\{\mathrm{v}(\mu) + \mathrm{v}'(\mu)(\mathrm{x} - \mu) + 1/2\mathrm{v}''(\mu)(\mathrm{x} - \mu)2\}$$

Thus,

$$\mathrm{v}(\mu) + \mathrm{v}'(\mu)(\mathrm{CE} - \mu) \approx \mathrm{E}\{\mathrm{v}(\mu) + \mathrm{v}'(\mu)(\mathrm{x} - \mu) + 1/2\mathrm{v}''(\mu)(\mathrm{x} - \mu)2\}$$

Canceling v(μ) from both sides and noting that the average value of x is μ, so E(x – μ) = 0, we have $\mathrm{v}'(\mu)(\mathrm{CE} - \mu) \approx 1/2\mathrm{v}''(\mu)\sigma 2$

Then, dividing by $\mathrm{v}'(\mathrm{x}), \mu - \mathrm{CE} \approx 12\mathrm{v}(\mu)\mathrm{v}'(\mu)\sigma 2 = 12\rho(\mu)\sigma 2$

That is, the risk premium—the difference between the average outcome and the certainty equivalent—is approximately equal to the Arrow-Pratt measure times half the variance, at least when the variance is small.

The translation of risk into dollars, by way of a risk premium, can be assessed even for large gambles if we are willing to make some technical assumptions. If a utility has **constant absolute risk aversion (CARA)**, the measure of risk aversion doesn't change with wealth; that is ρ=− v ″ (x) v ′ (x) is a constant. This turns out to imply, after setting the utility of zero to zero, that $v(x) = 1\rho(1 - e - \rho x)$. (This formulation is derived by setting v(0) = 0, handling the case of ρ = 0 with appropriate limits.) Now also assume that the gamble x is normally distributed with mean μ and variance δ^2. Then the expected value of v(x) is $\mathrm{Ev}(\mathrm{x}) = 1\rho(1 - \mathrm{e} - \rho(\mu - \rho 2\sigma 2))$

It is an immediate result from this formula that the certainty equivalent, with CARA preferences and normal risks, is μ− ρ 2 σ 2 . Hence, the risk premium of a normal distribution for a CARA individual is ρ 2 σ 2 . This formulation will appear when we consider agency theory and the challenges of motivating a risk averse employee when outcomes have a substantial random component.

An important aspect of CARA with normally distributed risks is that the preferences of the consumer are linear in the mean of the gamble and the variance. In fact, given a choice of gambles, the consumer selects the one with the highest value of μ− ρ 2 σ 2 . Such preferences are often called **mean variance preferences**, and they describe people who value risk linearly with the expected return. Such preferences comprise the foundation of modern finance theory.

Key Takeaways

- The von Neumann-Morgenstern utility function for a given person is a value v(x) for each possible outcome x, and the average value of v is the value the person assigns to the risky outcome. Under this theory, people value risk at the expected utility of the risk.
- The von Neumann approach is the prevailing model of behavior under risk, although there are numerous experiment-based criticisms of the theory.
- An individual dislikes risk if v is concave.
- For smoothly differentiable functions, concavity is equivalent to a second derivative that is not positive.
- People with concave von Neumann-Morgenstern utility functions are known as risk-averse people.
- The certainty equivalent of a gamble is an amount of money that provides equal utility to the random payoff of the gamble. The certainty equivalent is less than the expected outcome if the person is risk averse.
- The risk premium is defined to be the difference between the expected payoff and the certainty equivalent.
- The risk premium falls as wealth increases for any gamble, if and only if − v ″ (x) v ′ (x) is decreasing.
- The measure ρ(x)=− v ″ (x) v ′ (x) is known as the Arrow-Pratt measure of risk aversion, and also as the measure of absolute risk aversion.
- The risk premium is approximately equal to the Arrow-Pratt measure times half the variance when the variance is small.
- Constant absolute risk aversion provides a basis for “mean variance preferences,” the foundation of modern finance theory.

EXERCISES

1. Use a quadratic approximation on both sides of the equation to sharpen the estimate of the risk premium. First, note that $v(\mu) + v'(\mu)(CE - \mu) + ½v''(\mu)(CE - \mu)2 \approx v(CE) = Ev(x) \approx Ev(\mu) + v'(\mu)(x - \mu) + ½v''(\mu)(x - \mu)2.$
2. Conclude that $\mu - \text{CE} \approx 1\rho(1 + \rho 2\sigma 2 - 1)$. This approximation is exact to the second order.
3. Suppose that $u(x) = x^{0.95}$ for a consumer with a wealth level of $50,000. Consider a gamble, with equal probability of winning $100 and losing $100, and compute the risk premium associated with the gamble.
4. Suppose that $u(x) = x^{0.99}$ for a consumer with a wealth level of $100,000. A lottery ticket costs $1 and pays $5,000,000 with the probability 1 10,000,000 . Compute the certainty equivalent of the lottery ticket.
5. The return on U.S. government treasury investments is approximately 3%. Thus, a $1 investment returns $1.03 after one year. Treat this return as risk-free. The stock market (S&P 500) returns 7% on average and has a variance that is around 16% (the variance of return on a $1 investment is $0.16). Compute the value of ρ for a CARA individual. What is the risk premium associated with equal probabilities of a $100 gain or loss given the value of ρ?
6. A consumer has utility $u(x) = x^{7/8}$ and a base wealth of $100,000. She is about to take part in a gamble that will give her $10,000 (bringing her to $110,000) if a fair die rolls less than 3 (probability 1/3), but will cost her $5,000 (leaving her with $95,000) otherwise.
    1. What is the certainty equivalent of participating in this gamble?
    2. How much would she be willing to pay in order to avoid this gamble?

This page titled 13.4: Risk Aversion is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

13.5: Search 13.5: SearchEdit Update titlecancel

Learning Objectives

- How should a consumer go about finding the lowest price when available prices are random?

In most communities, grocery stores advertise sale prices every Wednesday in a newspaper insert, and these prices can vary from week to week and from store to store. The price of gasoline can vary as much as 15 cents per gallon in a one-mile radius. Should you decide that you want to buy a specific Sony television, you may see distinct prices at Best Buy and other electronics retailers. For many goods and services, there is substantial variation in prices, which implies that there are gains for buyers to search for the best price.

The theory of consumer search behavior is just a little bit arcane, but the basic insight will be intuitive enough. The general idea is that, from the perspective of a buyer, the price that is offered is random and has a probability density function f (p). If a consumer faces a cost of search (e.g., if you have to visit a store—in person, telephonically, or virtually—the cost includes your time and any other costs necessary to obtain a price quote), the consumer will set a **reservation price**, which is a maximum price that he or she will pay without visiting another store. That is, if a store offers a price below p*, the consumer will buy; otherwise, he or she will visit another store, hoping for a better price.

Call the reservation price p*, and suppose that the cost of search is c. Let J(p*) represent the expected total cost of purchase (including search costs). Then J must equal $J(p*) = c + \int 0p * pf(p)dp + \int p * \infty J(p*)f(p)dp$.

This equation arises because the current draw (which costs c) could either result in a price less than p*, in which case observed price, with density f, will determine the price paid p; or the price will be too high, in which case the consumer is going to take another draw, at cost c, and on average get the average price J(p*). It is useful to introduce the cumulative distribution function F, with $F(x) = \int 0 \times f(p)dp$. Note that something has to happen, so F(∞) = 1.

We can solve the equality for $J(p*), J(p*) = \int 0p * pf(p)dp + cF(p*)$.

This expression has a simple interpretation. The expected price J(p*) is composed of two terms. The first is the expected price, which is $\int 0\mathrm{p}^{*}\mathrm{pf(p)F\,(p^{*})\,dp}$. This has the interpretation of the average price conditional on that price being less than p*. This is because f(p) F(p*) is, in fact, the density of the random variable which is the price given that the price is less than p*. The second term is c F(p*) . This is the expected search cost, and it arises because 1 F(p*) is the expected number of searches. This arises because the odds of getting a price low enough to be acceptable is F(p*). There is a general statistical property underlying the number of searches. Consider a basketball player who successfully shoots a free throw with probability y. How many balls, on average, must he throw to sink one basket? The answer is 1/y. To see this, note that the probability that exactly n throws are required is $(1-y)^{n-1}$ y. This is because n are required means that n – 1 must fail (probability $(1 – y)^{n–1}$) and then the remaining one goes in, with probability y. Thus, the expected number of throws is

$$y + 2(1-y)y + 3(1-y)2y + 4(1-y)3y + ... = y(1 + 2(1-y) + 3(1-y)2 + 4(1-y)3 + ...) = y$$
$$((1 + (1-y) + (1-y)2 + (1-y)3 + ...) + (1-y)(1 + (1-y) + (1-y)2 + (1-y)3 + ...)$$
$$= (1-y)2(1 + (1-y) + (1-y)2 + (1-y)3 + ...) + (1-y)3(1 + (1-y) + (1-y)2 + ...) + ...$$
$$= y(1y + (1-y)1y + (1-y)21y + (1-y)31y + ...) = 1y.$$

Our problem has the same logic—where a successful basketball throw corresponds to finding a price less than p*.

The expected total cost of purchase, given a reservation price p*, is given by $J(p*) = \int 0p * pf(p)dp + cF(p*)$.

But what value of p* minimizes cost? Let's start by differentiating

$$J'(p*) = p * f(p*)F(p*) - f(p*) \int 0p * pf(p)dp + cF(p*)2$$
$$= f(p*)F(p*)(p * - \int 0p * pf(p)dp + cF(p*)) = f(p*)F(p*)(p * -J(p*)).$$

Thus, if p* < J(p*), J is decreasing, and it lowers cost to increase p*. Similarly, if p* > J(p*), J is increasing in p*, and it reduces cost to decrease p*. Thus, minimization occurs at a point where p* = J(p*).

Moreover, there is only one such solution to the equation p* = J(p*) in the range where f is positive. To see this, note that at any solution to the equation p* = J(p*), J ′ (p*)=0 and

$$J''(p*) = ddp * (f(p*)F(p*)(p * - J(p*))) = (ddp * f(p*)F(p*))(p * - J(p*)) + f(p*)F(p*)(1 - J'(p*)) = f(p*)F(p*) > 0$$

$J'(p^*) - \omega p * (J(p^*)F(p^*)(p* - J(p^*))) - (\omega p * J(p^*)F(p^*))(p* - J(p^*)) + J(p^*)F(p^*)(1 - J'(p^*)) - J(p^*)F(p^*) > 0$

This means that J takes a minimum at this value, since its first derivative is zero and its second derivative is positive, and that is true about any solution to p* = J(p*). Were there to be two such solutions, J ′ would have to be both positive and negative on the interval between them, since J is increasing to the right of the first (lower) one, and decreasing to the left of the second (higher) one. Consequently, the equation p* = J(p*) has a unique solution that minimizes the cost of purchase.

Consumer search to minimize cost dictates setting a reservation price equal to the expected total cost of purchasing the good, and purchasing whenever the price offered is lower than that level. That is, it is not sensible to “hold out” for a price lower than what you expect to pay on average, although this might be well useful in a bargaining context rather than in a store searching context.

Example (Uniform): Suppose that prices are uniformly distributed on the interval [a, b]. For p* in this interval,

$$J(p*) = \int 0p * pf(p)dp + cF(p*) = \int ap * pdpb - a + cp * -ab - a$$
$$= ½(p * 2 - a2) + c(b - a)p * -a = ½(p * +a) + c(b - a)p * -a.$$

Thus, the first-order condition for minimizing cost is $0 = J'(p^*) = 1/2 - c(b-a)(p^* - a)2$, implying $p^* = a + 2c(b-a)$.

There are a couple of interesting observations about this solution. First, not surprisingly, as c → 0, p* → α; that is, as the search costs go to zero, one holds out for the lowest possible price. This is sensible in the context of the model, but in real search situations delay may also have a cost that isn’t modeled here. Second, p* < b, the maximum price, if 2c < (b – a). In other words, if the most you can save by a search is twice the search cost, then don’t search, because the expected gains from the search will be half the maximum gains (thanks to the uniform distribution) and the search will be unprofitable.

The third observation, which is much more general than the specific uniform example, is that the expected price is a concave function of the cost of search (second derivative negative). That is, in fact, true for any distribution. To see this, define a function $H(c) = \min p^* J(p^*) = \min p^* \int 0p^* pf(p)dp + cF(p^*)$. Since $J'(p^*) = 0, H'(c) = \partial \partial cJ(p^*) = 1F(p^*)$

From here it requires only a modest effort to show that p* is increasing in c, from which it follows that H is concave. This means that there are increasing returns to decreasing search costs in that the expected total price of search is decreasing at an increasing rate as the cost of search decreases.

Key Takeaways

- For many goods, prices vary across location and across time. In response to price variation, consumers will often search for low prices.
- In many circumstances the best strategy is a reservation price strategy, where the consumer buys whenever offered a price below the reservation price.
- Consumer search to minimize cost dictates setting a reservation price equal to the expected total cost of purchasing the good, and purchasing whenever the price offered is lower than that level.

EXERCISE

1. Suppose that there are two possible prices, one and two, and that the probability of the lower price one is x. Compute the consumer’s reservation price, which is the expected cost of searching, as a function of x and the cost of search c. For what values of x and c should the consumer accept two on the first search or continue searching until the lower price one is found?

This page titled 13.5: Search is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

14: General Equilibrium

14.1: Edgeworth Box

14.2: Equilibrium With Price System

14.3: General Equilibrium

This page titled 14: General Equilibrium is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

14.1: Edgeworth Box 14.1: Edgeworth BoxEdit Update titlecancel

Learning Objectives

- How are several prices simultaneously determined?
- What are the efficient allocations?
- Does a price system equilibrium yield efficient prices?

The EdgeworthFrancis Edgeworth (1845–1926) introduced a variety of mathematical tools, including calculus, for considering economics and political issues, and was certainly among the first to use advanced mathematics for studying ethical problems. box considers a two-person, two-good "exchange economy." That is, two people have utility functions of two goods and endowments (initial allocations) of the two goods. The **Edgeworth box** is a graphical representation of the exchange problem facing these people and also permits a straightforward solution to their exchange problem.

Figure 14.1 The Edgeworth box

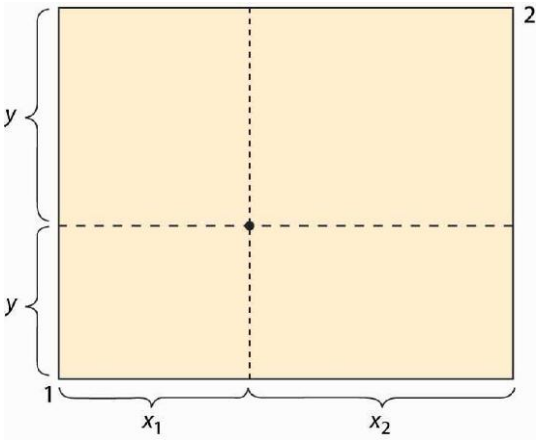


Figure 14.1. Person 1 is "located" in the lower left (southwest) corner, and Person 2 in the upper right (northeast) corner. The X good is given on the horizontal axis, the Y good on the vertical. The distance between them is the total amount of the good that they have between them. A point in the box gives the allocation of the good—the distance to the lower left to Person 1, the remainder to Person 2. Thus, for the point illustrated, Person 1 obtains (x_1, y_1), and Person 2 obtains (x_2, y_2). The total amount of each good available to the two people will be fixed.

What points are efficient? The economic notion of efficiency is that an allocation is efficient if it is impossible to make one person better off without harming the other person; that is, the only way to improve 1's utility is to harm 2, and vice versa. Otherwise, if the consumption is inefficient, there is a rearrangement that makes both parties better off, and the parties should prefer such a point. Now, there is no sense of fairness embedded in the notion, and there is an efficient point in which one person gets everything and the other gets nothing. That might be very unfair, but it could still be the case that improving 2 must necessarily harm 1. The allocation is efficient if there is no waste or slack in the system, even if it is wildly unfair. To distinguish this economic notion, it is sometimes called **Pareto efficiency**.Vilfredo Pareto (1848–1923) was a pioneer in replacing concepts of utility with abstract preferences. His work was later adopted by the economics profession and remains the modern approach.

We can find the Pareto-efficient points by fixing Person 1's utility and then asking what point, on the indifference isoquant of Person 1, maximizes Person 2's utility. At that point, any increase in Person 2's utility must come at the expense of Person 1, and vice versa; that is, the point is Pareto efficient. An example is illustrated in Figure 14.2.

Figure 14.2 An efficient point

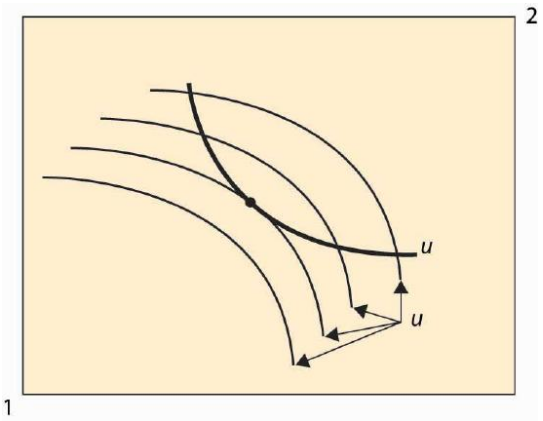


Figure 14.2, the isoquant of Person 1 is drawn with a dark, thick line. This utility level is fixed. It acts like the "budget constraint" for Person 2. Note that Person 2's isoquants face the opposite way because a movement southwest is good for 2, since it gives him more of both goods. Four isoquants are graphed for Person 2, and the highest feasible isoquant, which leaves Person 1 getting the fixed utility, has the Pareto-efficient point illustrated with a large dot. Such points occur at tangencies of the isoquants.

This process of identifying the points that are Pareto efficient can be carried out for every possible utility level for Person 1. What results is the set of Pareto-efficient points, and this set is also known as the **contract curve**. This is illustrated with the thick line in Figure 14.3. Every point on this curve maximizes one person's utility given the other's utility, and they are characterized by the tangencies in the isoquants.

The contract curve need not have a simple shape, as Figure 14.3 illustrates. The main properties are that it is increasing and ranges from Person 1 consuming zero of both goods to Person 2 consuming zero of both goods.

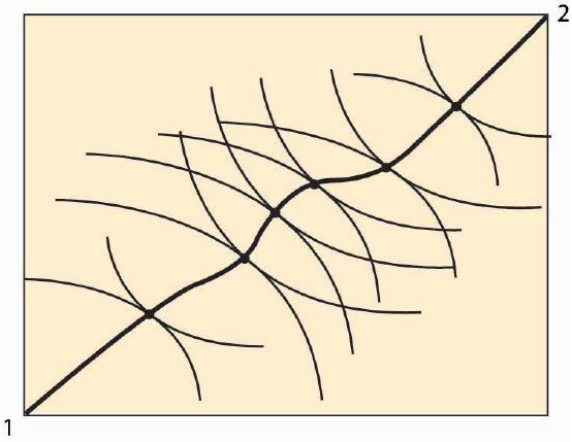


Figure 14.3 The contract curve

Example: Suppose that both people have Cobb-Douglas utility. Let the total endowment of each good be one, so that $x_2 = 1 - x_1$. Then Person 1's utility can be written as

$u_1 = x^a y^{1-a}$, and 2's utility is $u_2 = (1-x)^\beta (1-y)^{1-\beta}$ Then a point is Pareto efficient if

$$\alpha y(1-\alpha)x = \partial u1 \partial x \partial u1 \partial y = \partial u2 \partial x \partial u2 \partial y = \beta(1-y)(1-\beta)(1-x).$$

Thus, solving for y, a point is on the contract curve if

$$y = (1-\alpha)\beta x(1-\beta)\alpha + (\beta-\alpha)x = x(1-\beta)\alpha(1-\alpha)\beta + \beta - \alpha(1-\alpha)\beta x = xx + ((1-\beta)\alpha(1-\alpha)\beta)(1-x).$$

Thus, the contract curve for the Cobb-Douglas case depends on a single parameter $(1-\beta)\alpha\ (1-\alpha)\beta$. It is graphed for a variety of examples (α and β) in Figure 14.4.

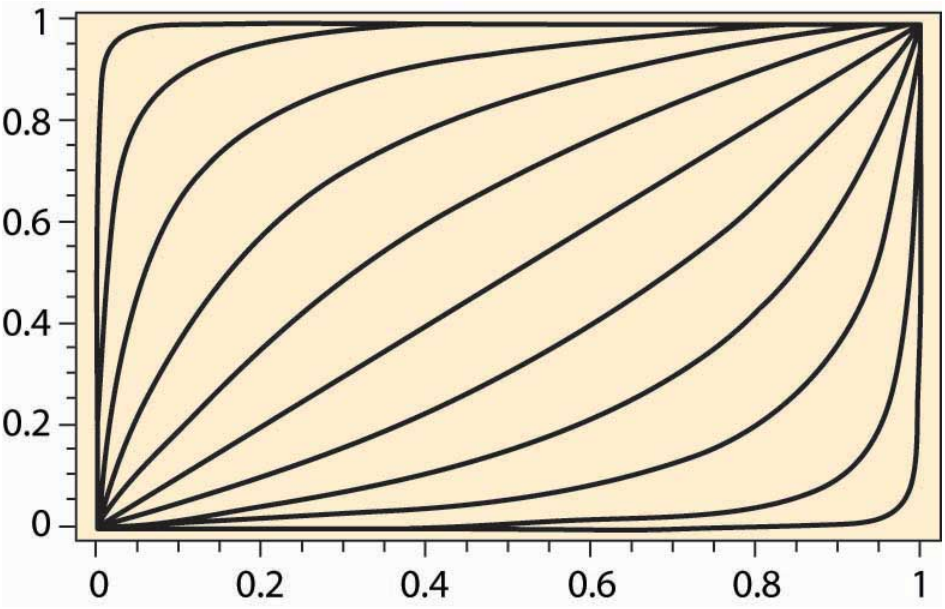


Figure 14.4 Contract curves with Cobb-Douglas utility

Key Takeaways

- The Edgeworth box considers a two-person, two-good "exchange economy." The Edgeworth box is a graphical representation of the exchange problem facing these people and also permits a straightforward solution to their exchange problem. A point in the Edgeworth box is the consumption of one individual, with the balance of the endowment going to the other.
- Pareto efficiency is an allocation in which making one person better off requires making someone else worse off—there are no gains from trade or reallocation.
- In the Edgeworth box, the Pareto-efficient points arise as tangents between isoquants of the individuals. The set of such points is called the contract curve. The contract curve is always increasing.

EXERCISES

1. If two individuals have the same utility function concerning goods, is the contract curve the diagonal line? Why or why not?
2. For two individuals with Cobb-Douglas preferences, when is the contract curve the diagonal line?

This page titled 14.1: Edgeworth Box is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

14.2: Equilibrium With Price System 14.2: Equilibrium With Price SystemEdit Update titlecancel

Learning Objectives

- How are prices in the two-person economy determined?
- Are these prices efficient?

The contract curve provides the set of efficient points. What point will actually be chosen? Let's start with an endowment of the goods. An endowment is just a point in the Edgeworth box that represents the initial ownership of both goods for both people. The endowment is marked with a triangle in Figure 14.5. Note this point indicates the endowment of both Person 1 and Person 2 because it shows the shares of each.

Figure 14.5 Individually rational efficient points

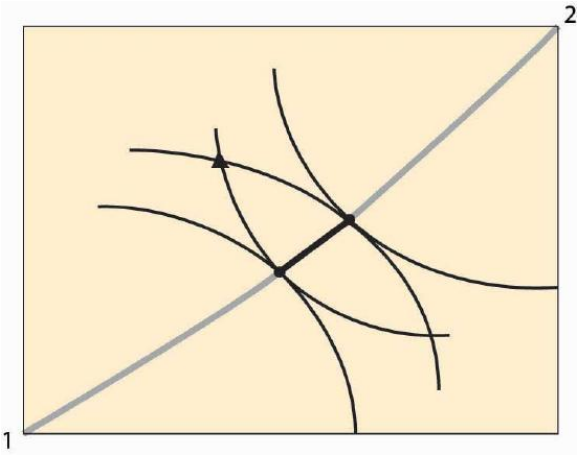


Figure 14.5 also shows isoquants for persons 1 and 2 going through the endowment. Note the isoquant for 1 is concave toward the point labeled 1, and the isoquant for 2 is concave toward the point labeled 2. These utility isoquants define a reservation utility level for each person—the utility they could get alone, without exchange. This "no exchange" state is known as **autarky**. There are a variety of efficient points that give these people at least as much as they get under autarky, and those points are along the contract curve but have a darker line.

In Figure 14.5, starting at the endowment, the utility of both players is increased by moving in a southeast direction—that is, down and to the right—until the contract curve is reached. This involves Person 1 getting more *X* (movement to the right) in exchange for giving up some *Y* (movement down). Thus, we can view the increase in utility as a trade—Person 1 trades some of his *Y* for some of Person 2's *X*.

In principle, any of the darker points on the contract curve, which give both people at least as much as they achieve under autarky, might result from trade. The two people get together and agree on exchange that puts them at any point along this segment of the curve, depending upon the bargaining skills of the players. But there is a particular point, or possibly a set of points, that results from exchange using a **price system**. A price system involves a specific price for trading *Y* for *X*, and vice versa, that is available to both parties. In this figure, prices define a straight line whose slope is the negative of the *Y* for *X* price (the *X* for *Y* price is the reciprocal).

Figure 14.6 Equilibrium with a price system

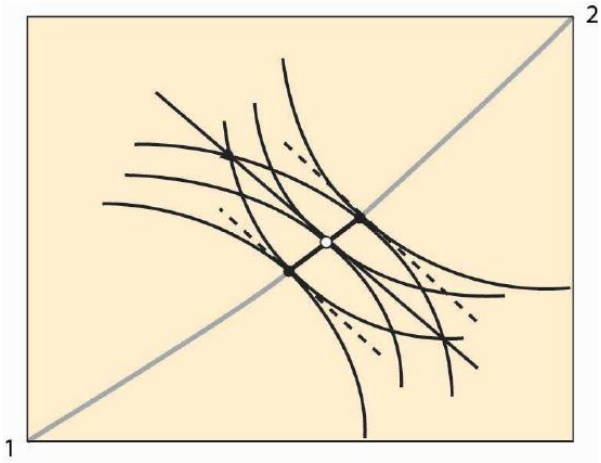


Figure 14.6 illustrates trade with a price system. The O in the center is the point on the contract curve connected to the endowment (triangle) by a straight line (the price line) in such a way that the straight line is tangent to both 1 and 2's isoquants at the contract curve. This construction means that if each person took the price line as a budget constraint, they would maximize their utility function by choosing the point labeled O.

The fact that a price line exists, that (i) goes through the endowment and (ii) goes through the contract curve at a point tangent to both people's utility, is relatively easy to show. Consider lines that satisfy property (ii), and let's see if we can find one that goes through the endowment. Start on the contract curve at the point that maximizes 1's utility given 2's reservation utility, and you can easily see that the price line through that point passes above and to the right of the endowment. The similar price line maximizing 2's utility given 1's reservation utility passes below and to the left of the endowment. These price lines are illustrated with dotted lines. Thus, by continuity, somewhere in between is a price line that passes through the endowment.

The point labeled O represents an equilibrium of the price system, in so far as supply and demand are equated for both goods. Note, given the endowment and the price through the endowment, both parties maximize utility by going to the O. To see this, it may help to consider a version of the figure that only shows Person 2's isoquants and the price line.

Figure 14.7 Illustration of price system equilibrium

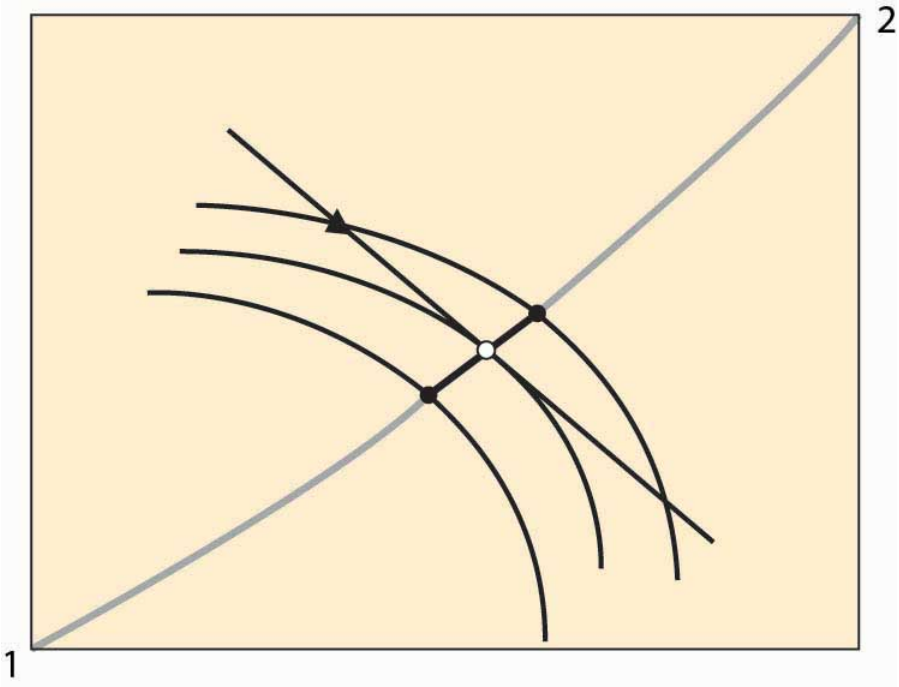


Figure 14.7 removes player 1's isoquants, leaving only player 2's isoquants and the price line through the endowment. The price line through the endowment is the budget facing each player at that price. Note that, given this budget line, player 2, who gets more as player 1 gets less, maximizes utility at the middle isoquant. That is, taking the price as given, player 2 would choose the O given player 2's endowment. The logic for player 1 is analogous. This shows that if both players believe that they can buy or sell as much as they like at the trade-off of the price through the O, both will trade to reach the O. This means that if the players accept the price, a balance of supply and demand emerges. In this sense, we have found an equilibrium price.

In the Edgeworth box, we see that, given an endowment, it is possible to reach some Pareto-efficient point using a price system. Moreover, any point on the contract curve arises as an equilibrium of the price system for some endowment. The proof of this proposition is startlingly easy. To show that a particular point on the contract curve is an equilibrium for some endowment, just start with an endowment equal to the point on the contract curve. No trade can occur because the starting point is Pareto efficient—any gain by one party entails a loss by the other.

Furthermore, if a point in the Edgeworth box represents an equilibrium using a price system (that is, if the quantity supplied equals the quantity demanded for both goods), it must be Pareto efficient. At an equilibrium to the price system, each player's isoquant is tangent to the price line and, hence, tangent to each other. This implies that the equilibrium is Pareto efficient.

Two of the three propositions are known as the first and second welfare theorems of general equilibrium. The **first welfare theorem of general equilibrium** states that any equilibrium of the price system is Pareto efficient. The **second welfare theorem of general equilibrium** states that any Pareto-efficient point is an equilibrium of the price system for some endowment. They have been demonstrated by Nobel laureates Kenneth Arrow and Gerard Debreu, for an arbitrary number of people and goods. They also demonstrated the third proposition—that, for any endowment, there exists an equilibrium of the price system with the same high level of generality.

Key Takeaways

- Autarky means consuming one's endowment without trade.
- If the endowment is not on the contract curve, there are points on the contract curve that make both people better off.
- A price system involves a specific price for trading *Y* for *X*, and vice versa, that is available to both parties. Prices define a straight line whose slope is the negative of the *Y* for *X* price (the *X* for *Y* price is the reciprocal).
- There is a price that (i) goes through the endowment and (ii) goes through the contract curve at a point tangent to both people's utility. Such a price represents a supply and demand equilibrium: Given the price, both parties would trade to the same point on the contract curve.
- In the Edgeworth box, it is possible to reach some Pareto-efficient point using a price system. Moreover, any point on the contract curve arises as an equilibrium of the price system for some endowment.
- If a point in the Edgeworth box represents an equilibrium using a price system, it must be Pareto efficient.

- The first and second welfare theorems of general equilibrium are that any equilibrium of the price system is Pareto efficient and any Pareto-efficient point is an equilibrium of the price system for some endowment.

This page titled 14.2: Equilibrium With Price System is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

14.3: General Equilibrium 14.3: General EquilibriumEdit Update titlecancel

Learning Objectives

- What happens in a general equilibrium when there are more than two people buying more than two goods?
- Does the Cobb-Douglas case provide insight?

We will illustrate general equilibrium for the case when all consumers have Cobb-Douglas utility in an **exchange economy**. An exchange economy is an economy where the supply of each good is just the total endowment of that good, and there is no production. Suppose that there are N people, indexed by $n = 1, 2, \ldots, N$ There are G goods, indexed by $g = 1, 2, \ldots, G$. Person n has Cobb-Douglas utility, which we can represent using exponents α(n, g), so that the utility of person n can be represented as $\Pi g = 1G \times (n, g)a(n, g)$, where x(n, g) is person n's consumption of good g. Assume that $a(n, g) \geq 0$ for all n and g, which amounts to assuming that the products are, in fact, goods. Without any loss of generality, we can require $\sum g = 1Ga(n, g) = 1$ for each n. (To see this, note that maximizing the function U is equivalent to maximizing the function U^{β} for any positive β.)

Let y(n, g) be person n's endowment of good g. The goal of general equilibrium is to find prices p_1, p_2, … , p_G for the goods in such a way that demand for each good exactly equals supply of the good. The supply of good g is just the sum of the endowments of that good. The prices yield a wealth for person n equal to W n = ∑ g=1 G p g y(n,g) .

We will assume that $\Sigma n = 1\mathrm{Na}(\mathrm{n}, \mathrm{g})\mathrm{y}(\mathrm{n}, \mathrm{i}) > 0$ for every pair of goods g and i. This assumption states that for any pair of goods, there is at least one agent that values good g and has an endowment of good i. The assumption ensures that there is always someone who is willing and able to trade if the price is sufficiently attractive. The assumption is much stronger than necessary but useful for exposition. The assumption also ensures that the endowment of each good is positive.

The Cobb-Douglas utility simplifies the analysis because of a feature that we already encountered in the case of two goods, which holds, in general, that the share of wealth for a consumer n on good g equals the exponent α(n, g). Thus, the total demand for good g is $\mathrm{Xg} = \Sigma \mathrm{n} = 1\mathrm{Na}(\mathrm{n}, \mathrm{g})\mathrm{Wnpg}$.

The equilibrium conditions, then, can be expressed by saying that supply (sum of the endowments) equals demand; or, for each good g, $\mathrm{Xg} = \Sigma \mathrm{n} = 1\mathrm{Na}(\mathrm{n}, \mathrm{g})\mathrm{Wnpg}$

We can rewrite this expression, provided that $p_g > 0$ (and it must be, for otherwise demand is infinite), to be

$$pg - \Sigma i = 1Gpi\Sigma n = 1Ny(n, i)a(n, g)\Sigma n = 1Ny(n, g) = 0$$

Let **B** be the G × G matrix whose (g, i) term is b $gi = \sum n = 1Ny(n, i)a(n, g) \sum n = 1Ny(n, g)$

Let **p** be the vector of prices. Then we can write the equilibrium conditions as **(I – B) p = 0**, where **0** is the zero vector. Thus, for an equilibrium (other than **p** = **0**) to exist, **B** must have an eigenvalue equal to 1 and a corresponding eigenvector **p** that is positive in each component. Moreover, if such an eigenvector–eigenvalue pair exists, it is an equilibrium, because demand is equal to supply for each good.

The actual price vector is not completely identified because if **p** is an equilibrium price vector, then so is any positive scalar times **p**. Scaling prices doesn't change the equilibrium because both prices and wealth (which is based on endowments) rise by the scalar factor. Usually economists assign one good to be a numeraire, which means that all other goods are indexed in terms of that good; and the numeraire's price is artificially set to be 1. We will treat any scaling of a price vector as the same vector.

The relevant theorem is the **Perron-Frobenius theorem**.Oskar Perron (1880–1975) and Georg Frobenius (1849–1917). It states that if **B** is a positive matrix (each component positive), then there is an eigenvalue λ > 0 and an associated positive eigenvector **p**; and, moreover, λ is the largest (in absolute value) eigenvector of **B**.The Perron-Frobenius theorem, as usually stated, only assumes that **B** is nonnegative and that **B** is irreducible. It turns out that a strictly positive matrix is irreducible, so this condition is sufficient to invoke the theorem. In addition, we can still apply the theorem even when **B** has some zeros in it, provided that it is irreducible. Irreducibility means that the economy can't be divided into two economies, where the people in one economy can't buy from the people in the second economy because they aren't endowed with anything that the people in the first economy value. If **B** is not irreducible, then some people may wind up consuming zero of things they value. This conclusion does most of the work of demonstrating the existence of an equilibrium. The only remaining condition to check is that the eigenvalue is in fact 1, so that **(I – B) p = 0**.

Suppose that the eigenvalue is λ. Then λ**p** = **Bp**. Thus for each g,

$$\lambda pg = \sum i = 1G \sum n = 1N\alpha(n,g)y(n,i) \sum m = 1Ny(m,g)pi$$

or

$$\lambda pg \sum m = 1Ny(m,g) = \sum i = 1G \sum n = 1N\alpha(n,g)y(n,i)pi.$$

Summing both sides over g,

$$\lambda \sum g = 1Gpg \sum m = 1Ny(m,g) = \sum g = 1G \sum i = 1G \sum n = 1N\alpha(n,g)y(n,i)pi = \sum i = 1G \sum n = 1N \sum g$$
$$= 1G\alpha(n,g)y(n,i)pi = \sum i = 1G \sum n = 1Ny(n,i)pi.$$

Thus, λ = 1 as desired.

The Perron-Frobenius theorem actually provides two more useful conclusions. First, the equilibrium is unique. This is a feature of the Cobb-Douglas utility and does not necessarily occur for other utility functions. Moreover, the equilibrium is readily approximated. Denote by $\mathbf{B}^t$ the product of **B** with itself t times. Then for any positive vector **v**, $\lim t \to \infty \mathrm{B}\ tv = p$. While approximations are very useful for large systems (large numbers of goods), the system can readily be computed exactly with small numbers of goods, even with a large number of individuals. Moreover, the approximation can be interpreted in a potentially useful manner. Let **v** be a candidate for an equilibrium price vector. Use **v** to permit people to calculate their wealth, which for person n is $\mathrm{Wn} = \Sigma \mathrm{i} = 1\mathrm{Gv\ i\ y(n,i)}$. Given the wealth levels, what prices clear the market? Demand for good g is $xg(v) = \sum n = 1N\alpha(n,g)Wn = \sum i = 1Gvi \sum n = 1N\alpha(n,g)y(n,i)$, and the market clears, given the wealth levels, if $pg = \sum i = 1Gvi \sum n = 1N\alpha(n,g)y(n,i) \sum n = 1Ny(n,g)$, which is equivalent to **p** = **Bv**. This defines an iterative process. Start with an arbitrary price vector, compute wealth levels, and then compute the price vector that clears the market for the given wealth levels. Use this price to recalculate the wealth levels, and then compute a new market-clearing price vector for the new wealth levels. This process can be iterated and, in fact, converges to the equilibrium price vector from any starting point.

We finish this section by considering three special cases. If there are two goods, we can let α_n = αα(n, 1), and then conclude that α(n, 2) = 1 – an. Then let Y g = ∑ n=1 N y(n,g) be the endowment of good g. Then the matrix **B** is

$$B = (1Y1 \sum n = 1Ny(n,1)an1Y1 \sum n = 1Ny(n,2)an1Y2 \sum n = 1Ny(n,1)(1-an)1Y2 \sum n = 1Ny(n,2)(1-an))$$
$$= (1Y1 \sum n = 1Ny(n,1)an1Y1 \sum n = 1Ny(n,2)an1Y2(Y1 - \sum n = 1Ny(n,1)an)1 - 1Y2 \sum n = 1Ny(n,2)an).$$

The relevant eigenvector of **B** is $p = (\sum n = 1Ny(n,2)an \sum n = 1Ny(n,1)(1-an))$.

The overall level of prices is not pinned down—any scalar multiple of **p** is also an equilibrium price—so the relevant term is the price ratio, which is the price of Good 1 in terms of Good 2, or

$$p1p2 = \sum n = 1Ny(n,2)an \sum n = 1Ny(n,1)(1-an).$$

We can readily see that an increase in the supply of Good 1, or a decrease in the supply of Good 2, decreases the price ratio. An increase in the preference for Good 1 increases the price of Good 1. When people who value Good 1 relatively highly are endowed with a lot of Good 2, the correlation between preference for Good 1, an, and endowment of Good 2 is higher. The higher the correlation, the higher is the price ratio. Intuitively, if the people who have a lot of Good 2 want a lot of Good 1, the price of Good 1 is going to be higher. Similarly, if the people who have a lot of Good 1 want a lot of Good 2, the price of Good 1 is going to be lower. Thus, the correlation between endowments and preferences also matters to the price ratio.

In our second special case, we consider people with the same preferences but who start with different endowments. Hypothesizing identical preferences sets aside the correlation between endowments and preferences found in the two-good case. Since people are the same, α(n, g) = Ag for all n. In this case, $bgi = \sum n = 1Ny(n,i)\alpha(n,g) \sum n = 1Ny(n,g) = AgYiYg$, whereas before $Yg = \sum n = 1Ny(n,g)$ is the total endowment of good g. The matrix **B** has a special structure, and in this case, $pg = AgYg$ is the equilibrium price vector. Prices are proportional to the preference for the good divided by the total endowment for that good.

Now consider a third special case, where no common structure is imposed on preferences, but endowments are proportional to each other; that is, the endowment of person n is a fraction wn of the total endowment. This implies that we can write $\mathrm{y(n,g)} = \mathrm{w_n Y_g}$, an equation assumed to hold for all people n and goods g. Note that by construction, $\sum n = 1Nwn = 1$, since the value wn represents n's share of the total endowment. In this case, we have

$$bgi = \sum n = 1Ny(n,i)\alpha(n,g) \sum n = 1Ny(n,g) = YiYg \sum n = 1Nwn\alpha(n,g)$$

These matrices also have a special structure, and it is readily verified that the equilibrium price vector satisfies $pg = 1Yg\sum n = 1Nwn\alpha(n,g)$.

This formula receives a similar interpretation—the price of good g is the strength of preference for good g, where strength of preference is a wealth-weighted average of the individual preference, divided by the endowment of the good. Such an interpretation is guaranteed by the assumption of Cobb-Douglas preferences, since these imply that individuals spend a constant proportion of their wealth on each good. It also generalizes the conclusion found in the two-good case to more goods, but with the restriction that the correlation is now between wealth and preferences. The special case has the virtue that individual wealth, which is endogenous because it depends on prices, can be readily determined.

Key Takeaways

- General equilibrium puts together consumer choice and producer theory to find sets of prices that clear many markets.
- For the case of an arbitrary number of goods and an arbitrary number of consumers—each with Cobb-Douglas utility—there is a closed form for the demand curves, and the price vector can be found by locating an eigenvector of a particular matrix. The equilibrium is unique (true for Cobb-Douglas but not true more generally).
- The actual price vector is not completely identified because if **p** is an equilibrium price vector, then so is any positive scalar times **p**. Scaling prices doesn't change the equilibrium because both prices and wealth (which is based on endowments) rise by the scalar factor.
- The intuition arising from one-good models may fail because of interactions with other markets—increasing preferences for a good (shifting out demand) changes the values of endowments in ways that then reverberate through the system.

EXERCISE

1. Consider a consumer with Cobb-Douglas utility $U = \prod i = 1Gxiai$,where $\Sigma i = 1\text{Gai} = 1$ and facing the budget constraint $W = \sum i = 1Gpixi$.Show that the consumer maximizes utility by choosing $xi = aiWpi$ for each good i. (Hint: Express the budget constraint as $xG = 1pG(W - \sum i = 1G - 1pixi)$,and thus utility as $U = (\prod i = 1G - 1xiai)(1pG(W - \sum i = 1G - 1pixi))aG.)$ This function can now be maximized in an unconstrained fashion. Verify that the result of the maximization can be expressed as $pixi = aiaGpGxG$, and thus $W = \sum i = 1Gpixi = \sum i = 1GaiaGpGxG = pGxGaG$,which yields $pixi = aiW$

This page titled 14.3: General Equilibrium is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

15: Monopoly

15.1: Sources of Monopoly
15.2: Basic Analysis
15.3: Effect of Taxes
15.4: Price Discrimination
15.5: Welfare Effects
15.6: Natural Monopoly
15.7: Peak-load Pricing

This page titled 15: Monopoly is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

15.1: Sources of Monopoly 15.1: Sources of MonopolyEdit Update titlecancel

Learning Objectives

- How do monopolies come about?

A **monopoly** is a firm that faces a downward sloping demand and has a choice about what price to charge—without fearing of chasing all of its customers away to rivals.

There are very few pure monopolies. The U.S. post office has a monopoly in first-class mail but faces competition from FedEx and other express-mail companies, as well as from fax and e-mail providers. Microsoft has a great deal of market power, but a small percentage of personal computer users choose Apple or Linux operating systems. In contrast, there is only one U.S. manufacturer of aircraft carriers.

However, many firms have **market power** or monopoly power, which means that they can increase their price above marginal cost and sustain sales for a long period of time.These terms are used somewhat differently among authors. Both terms require downward sloping demand and usually some notion of sustainability of sales. Some distinguish the terms by whether they are "large" or not; others by how long the price increase can be sustained. We won't need such distinctions here. The theory of monopoly is applicable to such firms, although they may face an additional and important constraint: A price increase may affect the behavior of rivals. This behavior of rivals is the subject of the next chapter.

A large market share is not proof of a monopoly, nor is a small market share proof that a firm lacks monopoly power. For example, U.S. Air dominated air traffic to Philadelphia and Pittsburgh but still lost money. Porsche has a small share of the automobile market—or even the high-end automobile market—but still has monopoly power in that market.

There are three basic sources of monopoly. The most common source is to be granted a monopoly by the government, either through patents—in which case the monopoly is temporary—or through a government franchise. Intelsat was a government franchise that was granted a monopoly on satellite communications, a monopoly that ultimately proved quite lucrative. Many cities and towns license a single cable TV company or taxi company, although usually basic rates and fares are set by the terms of the license agreement. New drugs are granted patents that provide the firms monopoly power for a period of time. (Patents generally last 20 years, but pharmaceutical drugs have their own patent laws.) Copyright also confers a limited monopoly for a limited period of time. Thus, the Disney Corporation owns copyrights on Mickey Mouse—copyrights that, by law, should have expired but were granted an extension by Congress each time they were due to expire. Copyrights create monopoly power over music as well as cartoon characters. Time Warner owns the rights to the song "Happy Birthday to You" and receives royalties every time that it is played on the radio or other commercial venue.Fair-use provisions protect individuals with noncommercial uses of copyrighted materials. Many of the Beatles' songs that Paul McCartney coauthored were purchased by Michael Jackson.

A second source of monopoly is a large economy of scale. The scale economy needs to be large relative to the size of demand. A monopoly can result when the average cost of a single firm serving the entire market is lower than that of two or more firms serving the market. For example, long-distance telephone lines were expensive to install, and the first company to do so, AT&T, wound up being the only provider of long-distance service in the United States. Similarly, scale economies in electricity generation meant that most communities had a single electricity provider prior to the 1980s, when new technology made relatively smaller scale generation more efficient.

The demand-side equivalent of an economy of scale is a **network externality**. A network externality arises when others' use of a product makes it more valuable to each consumer. Standards are a common source of network externality. Since AA batteries are standardized, it makes them more readily accessible, helps drive down their price through competition and economies of scale, and thus makes the AA battery more valuable. They are available everywhere, unlike proprietary batteries. Fax machines are valuable only if others have similar machines. In addition to standards, a source of network externality is third-party products. Choosing Microsoft Windows as a computer operating system means that there is more software available than for Macintosh or Linux, as the widespread adoption of Windows has led to the writing of a large variety of software for it. The JVC Video Home System of VCRs came to dominate the Sony Beta system, primarily because there were more movies to rent in the VHS format than in the Beta format at the video rental store. In contrast, recordable DVDs have been hobbled by incompatible standards of DVD+R and DVD-R, a conflict not resolved even as the next generation—50 GB discs such as Sony's Blu-ray—starts to reach the market. DVDs themselves were slow to be adopted by consumers because few discs were available for rent at video rental stores, which is

a consequence of few adoptions of DVD players. As DVD players became more prevalent and the number of discs for rent increased, the market tipped and DVDs came to dominate VHS.

The third source of monopoly is control of an essential, or a sufficiently valuable, input to the production process. Such an input could be technology that confers a cost advantage. For example, software is run by a computer operating system and needs to be designed to work well with the operating system. There have been a series of allegations that Microsoft kept secret some of the "application program interfaces" used by Word as a means of hobbling rivals. If so, access to the design of the operating system itself is an important input.

Key Takeaways

- A monopoly is a firm that faces a downward sloping demand and has a choice about what price to charge—an increase in price doesn't send most, or all, of the customers away to rivals.
- There are very few pure monopolies. There are many firms that have market power or monopoly power, which means that they can increase their price above marginal cost and sustain sales for a long period of time.
- A large market share is not proof of a monopoly, nor is a small market share proof that a firm lacks monopoly power. There are three basic sources of monopoly: one created by government, like patents; a large economy of scale or a network externality; and control of an essential, or a sufficiently valuable, input to the production process.

This page titled 15.1: Sources of Monopoly is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

15.2: Basic Analysis 15.2: Basic AnalysisEdit Update titlecancel

Learning Objectives

- What are the basic effects of monopoly, compared to a competitive industry?

Even a monopoly is constrained by demand. A monopoly would like to sell lots of units at a very high price, but a higher price necessarily leads to a loss in sales. So how does a monopoly choose its price and quantity?

A monopoly can choose price, or a monopoly can choose quantity and let the demand dictate the price. It is slightly more convenient to formulate the theory in terms of quantity rather than price, because costs are a function of quantity. Thus, we let p(q) be the demand price associated with quantity q, and c(q) be the cost of producing q. The monopoly's profits are $\pi = p(q)q - c(q)$.

The monopoly earns the revenue pq and pays the cost c. This leads to the first-order condition for the profit-maximizing quantity

$$qm : 0 = \partial\pi\partial q = p(qm) + qmp'(qm) - c'(qm).$$

The term p(q)+q p ′ (q) is known as **marginal revenue**. It is the derivative of revenue pq with respect to quantity. Thus, a monopoly chooses a quantity qm where marginal revenue equals marginal cost, and charges the maximum price p(qm) that the market will bear at that quantity. Marginal revenue is below demand p(q) because demand is downward sloping. That is, $\mathrm{p(q)} + \mathrm{qp'(q)} < \mathrm{p(q)}$

Figure 15.1 Basic monopoly diagram

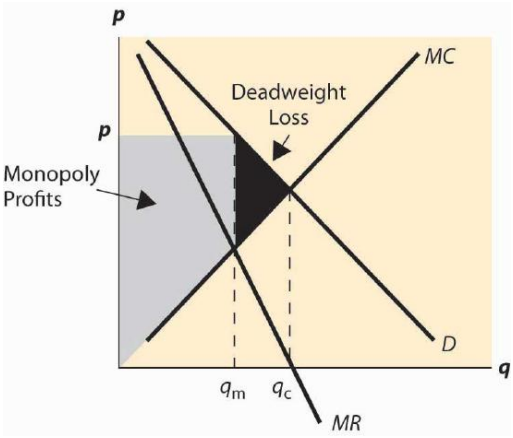


Figure 15.1. The key points of this diagram are fivefold. First, marginal revenue lies below the demand curve. This occurs because marginal revenue is the demand p(q) plus a negative number. Second, the monopoly quantity equates marginal revenue and marginal cost, but the monopoly price is higher than the marginal cost. Third, there is a deadweight loss, for the same reason that taxes create a deadweight loss: The higher price of the monopoly prevents some units from being traded that are valued more highly than they cost. Fourth, the monopoly profits from the increase in price, and the monopoly profit is shaded. Fifth, since—under competitive conditions—supply equals marginal cost, the intersection of marginal cost and demand corresponds to the competitive outcome. We see that the monopoly restricts output and charges a higher price than would prevail under competition.

We can rearrange the monopoly pricing formula to produce an additional insight:

$$p(qm) - c'(qm) = -qmp'(qm) \text{ or } p(qm) - c'(qm)p(qm) = -qmp'(qm)p(qm) = 1\varepsilon$$

The left-hand side of this equation (price minus marginal cost divided by price) is known as the **price-cost margin** or Lerner Index.Abba Lerner (1903–1982). Note that $1 - qmp'(qm)p(qm) = -1qmp'(qm)p(qm) = -dqqdpp = \varepsilon$m, which is used in the derivation. The right-hand side is one divided by the elasticity of demand. This formula relates the markup over marginal cost to the elasticity of demand. It is important because perfect competition forces price to equal marginal cost, so this formula provides a measure of the deviation from competition and, in particular, says that the deviation from competition is small when the elasticity of demand is large, and vice versa.

Marginal cost will always be greater than or equal to zero. If marginal cost is less than zero, the least expensive way to produce a given quantity is to produce more and throw some away. Thus, the price-cost margin is no greater than one; and, as a result, a monopolist produces in the elastic portion of demand. One implication of this observation is that if demand is everywhere inelastic

(e.g., $p(q) = q - a$ for $a > 1$), the optimal monopoly quantity is essentially zero, and in any event would be no more than one molecule of the product.

In addition, the effects of monopoly are related to the elasticity of demand. If demand is very elastic, the effect of monopoly on prices is quite limited. In contrast, if the demand is relatively inelastic, monopolies will increase prices by a large margin.

We can rewrite the formula to obtain $\mathrm{p(qm)} = \varepsilon\varepsilon - 1\mathrm{c}'(\mathrm{qm})$.

Thus, a monopolist marks up marginal cost by the factor ε ε−1 , at least when ε > 1. This formula is sometimes used to justify a "fixed markup policy," which means that a company adds a constant percentage markup to its products. This is an ill-advised policy, not justified by the formula, because the formula suggests a markup that depends upon the demand for the product in question, and thus not a fixed markup for all products that a company produces.

Key Takeaways

- Even a monopoly is constrained by demand.
- A monopoly can either choose price, or choose quantity and let the demand dictate the price.
- A monopoly chooses a quantity qm where marginal revenue equals marginal cost, and charges the maximum price p(qm) that the market will bear at that quantity.
- Marginal revenue is below demand p(q) because demand is downward sloping.
- The monopoly price is higher than the marginal cost.
- There is a deadweight loss of monopoly for the same reason that taxes create a deadweight loss: The higher price of the monopoly prevents some units from being traded that are valued more highly than they cost.
- A monopoly restricts output and charges a higher price than would prevail under competition.
- The price-cost margin is the ratio of price minus marginal cost over price and measures the deviation from marginal cost pricing.
- A monopoly chooses a price or quantity that equates the price-cost margin to the inverse of the demand elasticity.
- A monopolist produces in the elastic portion of demand.
- A monopolist marks up marginal cost by the factor ε ε−1 , when the elasticity of demand ε exceeds one.

EXERCISES

1. If demand is linear, $p(q) = a - bq$, what is marginal revenue? Plot demand and marginal revenue, and total revenue qp(q) as a function of q.
2. For the case of constant elasticity of demand, what is marginal revenue?
3. If both demand and supply have constant elasticity, compute the monopoly quantity and price.
4. Consider a monopolist with cost c = 3q.
    1. If demand is given by $q = 50 - 2p$, what is the monopoly price and quantity? What are the profits?
    2. Repeat part (a) for demand given by $q = 10/p$
5. The government wishes to impose a tax, of fraction t, on the profits of a monopolist. How does this affect the monopolist's optimal output quantity?
6. If demand has constant elasticity, what is the marginal revenue of the monopolist?

This page titled 15.2: Basic Analysis is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

15.3: Effect of Taxes 15.3: Effect of TaxesEdit Update titlecancel

Learning Objectives

- How does a monopoly respond to taxes?

A tax imposed on a seller with monopoly power performs differently than a tax imposed on a competitive industry. Ultimately, a perfectly competitive industry must pass on all of a tax to consumers because, in the long run, the competitive industry earns zero profits. In contrast, a monopolist might absorb some portion of a tax even in the long run.

To model the effect of taxes on a monopoly, consider a monopolist who faces a tax rate *t* per unit of sales. This monopolist earns $\pi = p(q)q - c(q) - tq.\backslash)$

The first-order condition for profit maximization yields $0 = \partial\pi\partial q = p(qm) + qmp'(qm) - c'(qm) - t.$

Viewing the monopoly quantity as a function of *t*, we obtain $dqmdt = 12p'(qm) + qmp''(qm) - c''(qm) < 0$ with the sign following from the second-order condition for profit maximization. In addition, the change in price satisfies $p'(qm)dqmdt = p'(qm)2p'(qm) + qmp''(qm) - c''(qm) > 0$

Thus, a tax causes a monopoly to increase its price. In addition, the monopoly price rises by less than the tax if $\mathrm{p'(qm)dqmdt} < 1, \text{ or } \mathrm{p'(qm) + qmp''(qm) - c''(qm) < 0}$

This condition need not be true but is a standard regularity condition imposed by assumption. It is true for linear demand and increasing marginal cost. It is false for constant elasticity of demand, $\varepsilon > 1$ (which is the relevant case, for otherwise the second-order conditions fail), and constant marginal cost. In the latter case (constant elasticity and marginal cost), a tax on a monopoly increases price by more than the amount of the tax.

Key Takeaways

- A perfectly competitive industry must pass on all of a tax to consumers because, in the long run, the competitive industry earns zero profits. A monopolist might absorb some portion of a tax even in the long run.
- A tax causes a monopoly to increase its price and reduce its quantity.
- A tax may or may not increase the monopoly markup.

EXERCISES

1. Use a revealed preference argument to show that a per-unit tax imposed on a monopoly causes the quantity to fall. That is, hypothesize quantities *qb* before the tax and *qa* after the tax, and show that two facts—the before-tax monopoly preferred *qb* to *qa*, and the taxed monopoly made higher profits from *qb*—together imply that $qb \leq qa$.
2. When both demand and supply have constant elasticity, use the results of 0 to compute the effect of a proportional tax (i.e., a portion of the price paid to the government).

This page titled 15.3: Effect of Taxes is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

15.4: Price Discrimination 15.4: Price DiscriminationEdit Update titlecancel

Learning Objectives

- Do monopolies charge different consumers different prices?
- Why and how much?

Pharmaceutical drugs for sale in Mexico are generally priced substantially below their U.S. counterparts. Pharmaceutical drugs in Europe are also cheaper than in the United States, although not as inexpensive as in Mexico, with Canadian prices usually falling between the U.S. and European prices. (The comparison is between identical drugs produced by the same manufacturer.)

Pharmaceutical drugs differ in price from country to country primarily because demand conditions vary. The formula p(q m)= ε ε−1 c ′ (q m) shows that a monopoly seller would like to charge a higher markup over marginal cost to customers with less elastic demand than to customers with more elastic demand because ε ε−1 is a decreasing function of ε, for $\varepsilon > 1$. Charging different prices for the same product to different customers is known as **price discrimination**. In business settings, it is sometimes known as value-based pricing, which is a more palatable term to relay to customers.

Computer software vendors often sell a “student” version of their software, usually at substantially reduced prices, but require proof of student status to qualify for the lower price. Such student discounts are examples of price discrimination, and students have more elastic demand than business users. Similarly, the student and senior citizen discounts at movies and other venues sell the same thing—a ticket to the show—for different prices, and thus qualify as price discrimination.

In order for a seller to price discriminate, the seller must be able to

- identify (approximately) the demand of groups of customers, and
- prevent arbitrage.

Arbitrage is also known as “buying low and selling high,” and represents the act of being an intermediary. Since price discrimination requires charging one group a higher price than another, there is potentially an opportunity for arbitrage, arising from members of the low-price group buying at the low price and selling at the high price. If the seller can’t prevent arbitrage, arbitrage essentially converts a two-price system to sales at the low price.

Why offer student discounts at the movies? You already know the answer to this: Students have lower incomes on average than others, and lower incomes translate into a lower willingness to pay for normal goods. Consequently, a discount to a student makes sense from a demand perspective. It is relatively simple to prevent arbitrage by requiring that a student identification card be presented. Senior citizen discounts are a bit subtler. Generally, senior citizens aren’t poorer than other groups of customers (in the United States, at least). However, seniors have more free time and therefore are able to substitute to matinee showingsMatinee showings are those early in the day, which are usually discounted. These discounts are not price discrimination because a show at noon isn’t the same product as a show in the evening. or to drive to more distant locations should those offer discounts. Thus, seniors have relatively elastic demand, more because of their ability to substitute than because of their income.

Airlines commonly price discriminate, using “Saturday night stay-overs” and other devices. To see that such charges represent price discrimination, consider a passenger who lives in Dallas but needs to spend Monday through Thursday in Los Angeles for 2 weeks in a row. This passenger could buy two roundtrip tickets:

Trip 1:

First Monday: Dallas → Los Angeles

First Friday: Los Angeles → Dallas

Trip 2:

Second Monday: Dallas → Los Angeles

Second Friday: Los Angeles → Dallas

At the time of this writing, the approximate combined cost of these two flights was $2,000. In contrast, another way of arranging exactly the same travel is to have two roundtrips, one of which originates in Dallas, while the other originates in Los Angeles:

Trip 1:

First Monday: Dallas → Los Angeles

Second Friday: Los Angeles → Dallas

Trip 2:

First Friday: Los Angeles → Dallas

Second Monday: Dallas → Los Angeles

This pair of roundtrips involves exactly the same travel as the first pair, but costs less than $500 for both (at the time of this writing). The difference is that the second pair involves staying over Saturday night for both legs, and that leads to a major discount for most U.S. airlines. (American Airlines quoted the fares.)

How can airlines price discriminate? There are two major groups of customers: business travelers and leisure travelers. Business travelers have the higher willingness to pay overall, and the nature of their trips tends to be that they come home for the weekend. In contrast, leisure travelers usually want to be away for a weekend, so a weekend stay over is an indicator of a leisure traveler. It doesn't work perfectly as an indicator—some business travelers must be away for the weekend—but it is sufficiently correlated with leisure travel that it is profitable for the airlines to price discriminate.

These examples illustrate an important distinction. Senior citizen and student discounts are based on the identity of the buyer, and qualifying for the discount requires that one show an identity card. In contrast, airline price discrimination is not based on the identity of the buyer but rather on the choices made by the buyer. Charging customers based on identity is known as **direct price discrimination**, while offering a menu or set of prices and permitting customers to choose distinct prices is known as **indirect price discrimination**.The older and incoherent language for these concepts identified direct price discrimination as "third-degree price discrimination," while indirect price discrimination was called second-degree price discrimination. In the older language, first-degree price discrimination meant perfect third-degree price discrimination.

Two common examples of indirect price discrimination are coupons and quantity discounts. Coupons offer discounts for products and are especially common in grocery stores, where they are usually provided in a free newspaper section at the front of the store. Coupons discriminate on the basis of the cost of time. It takes time to find the coupons for the products that one is interested in buying. Thus, those with a high value of time won't find it worth their while to spend 20 minutes to save $5 (effectively a $15 per hour return), while those with a low value of time will find that return worthwhile. Since those with a low value of time tend to be more price-sensitive (more elastic demand), coupons offer a discount that is available to all but used primarily by customers with a more elastic demand, and thus increase the profits of the seller.

Quantity discounts are discounts for buying more. Thus, the large size of milk, laundry detergent, and other items often cost less per unit than smaller sizes, and the difference is greater than the savings on packaging costs. In some cases, the larger sizes entail greater packaging costs; some manufacturers "band together" individual units, incurring additional costs to create a larger size that is then discounted. Thus, the "24-pack" of paper towels sells for less per roll than the individual rolls; such large volumes appeal primarily to large families, who are more price-sensitive on average.

Key Takeaways

- A monopoly seller would like to charge a higher markup over marginal cost to customers with less elastic demand than to customers with more elastic demand.
- In order for a seller to price discriminate, the seller must be able to
  - identify (approximately) the demand of groups of customers, and
  - prevent arbitrage.
- Since price discrimination requires charging one group a higher price than another, there is potentially an opportunity for arbitrage.
- Airlines commonly price discriminate, using "Saturday night stay-overs" and other devices.
- Direct price discrimination is based upon the identity of the buyer, while indirect price discrimination involves several offers and achieves price discrimination through customer choices.
- Two common examples of indirect price discrimination are coupons and quantity discounts.

EXERCISE

1. Determine whether the following items are direct price discrimination, indirect price discrimination, or not price discrimination—and why.
    1. Student discounts at local restaurants
    2. Financial aid at colleges
    3. Matinee discounts at the movies
    4. Home and professional versions of Microsoft’s operating system
    5. Lower airline fares for weekend flights
    6. Buy-one-get-one-free specials

This page titled 15.4: Price Discrimination is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

15.5: Welfare Effects 15.5: Welfare EffectsEdit Update titlecancel

Learning Objectives

- Is price discrimination good or bad for society as a whole?

Is price discrimination a good thing or a bad thing? It turns out that there is no definitive answer to this question. Instead, it depends on circumstances. We illustrate this conclusion with a pair of exercises.

This exercise illustrates a much more general proposition: If a price discriminating monopolist produces less than a nondiscriminating monopolist, then price discrimination reduces welfare. This proposition has an elementary proof. Consider the price discriminating monopolist's sales, and then allow arbitrage. The arbitrage increases the gains from trade, since every transaction has gains from trade. Arbitrage, however, leads to a common price like that charged by a nondiscriminating monopolist. Thus, the only way that price discrimination can increase welfare is if it leads a seller to sell more output than he or she would otherwise. This is possible, as the next exercise shows.

In Exercise 2, we see that price discrimination that brings in a new group of customers may increase the gains from trade. Indeed, this example involves a Pareto improvement: The seller and Group 2 are better off, and Group 1 is no worse off, than without price discrimination. (A Pareto improvement requires that no one is worse off and at least one person is better off.)

Whether price discrimination increases the gains from trade overall depends on circumstances. However, it is worth remembering that people with lower incomes tend to have more elastic demand, and thus get lower prices under price discrimination than absent price discrimination. Consequently, a ban on price discrimination tends to hurt the poor and benefit the rich, no matter what the overall effect.

A common form of price discrimination is known as **two-part pricing**. Two-part pricing usually involves a fixed charge and a marginal charge, and thus offers the ability for a seller to capture a portion of the consumer surplus. For example, electricity often comes with a fixed price per month and then a price per kilowatt-hour, which is two-part pricing. Similarly, long distance and cellular telephone companies charge a fixed fee per month, with a fixed number of "included" minutes, and a price per minute for additional minutes. Such contracts really involve three parts rather than two parts, but are similar in spirit.

Figure 15.2 Two-part pricing

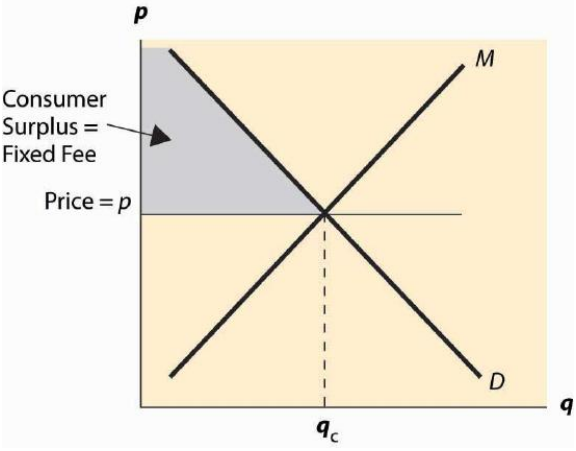


Figure 15.2.

Key Takeaways

- If a price discriminating monopolist produces less than a nondiscriminating monopolist, then price discrimination reduces welfare.
- Price discrimination that opens a new, previously unserved market increases welfare.
- A ban on price discrimination tends to hurt the poor and benefit the rich, no matter what the overall effect.
- Two-part pricing involves a fixed charge and a marginal charge.
- The ideal two-part price is to charge marginal cost plus a fixed charge equal to the customer's consumer surplus, in which case the seller captures the entire gains from trade.

EXERCISES

1. Let marginal cost be zero for all quantities. Suppose that there are two equal-sized groups of customers, Group 1 with demand $q(p) = 12 - p$, and Group 2 with demand $q(p) = 8–p$. Show that a nondiscriminating monopolist charges a price of 5, and the discriminating monopolist charges Group 1 the price of 6 and Group 2 the price of 4. Then calculate the gains from trade, with discrimination and without, and show that price discrimination reduces the gains from trade.
2. Let marginal cost be zero for all quantities. Suppose that there are two equal-sized groups of customers, Group 1 with demand $q(p) = 12–p$, and Group 2 with demand $q(p) = 4–p$. Show that a nondiscriminating monopolist charges a price of 6, and the discriminating monopolist charges Group 1 the price of 6 and Group 2 the price of 2. Then calculate the gains from trade, with discrimination and without, and show that price discrimination increases the gains from trade.

This page titled 15.5: Welfare Effects is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

15.6: Natural Monopoly 15.6: Natural MonopolyEdit Update titlecancel

Learning Objectives

- When there is a scale economy, what market prices will arise?
- How is the monopoly price constrained by the threat of entry?

A **natural monopoly** arises when a single firm can efficiently serve the entire market because average costs are lower with one firm than with two firms. An example is illustrated in Figure 15.3. In this case, the average total cost of a single firm is lower than if two firms were to split the output between them. The monopolist would like to price at *pm*, which maximizes profits.The monopoly price may or may not be *sustainable*. A monopoly price is not sustainable if it were to lead to entry, thereby undercutting the monopoly. The feasibility of entry, in turn, depends on whether the costs of entering are not recoverable ("*sunk*") and how rapidly entry can occur. If the monopoly price is not sustainable, the monopoly may engage in *limit pricing*, which is jargon for pricing to deter (limit) entry.

Historically, the United States and other nations have regulated natural monopoly products and supplies such as electricity, telephony, and water service. An immediate problem with regulation is that the efficient price—that is, the price that maximizes the gains from trade—requires a subsidy from outside the industry. We see the need for a subsidy in Figure 15.3 because the price that maximizes the gains from trade is p_1, which sets the demand (marginal value) equal to the marginal cost. At this price, however, the average total cost exceeds the price, so that a firm with such a regulated price would lose money. There are two alternatives. The product could be subsidized: Subsidies are used with postal and passenger rail services in the United States and historically for many more products in Canada and Europe, including airlines and airplane manufacture. Alternatively, regulation could be imposed to limit the price to p_2, the lowest break-even price. This is the more common strategy used in the United States.

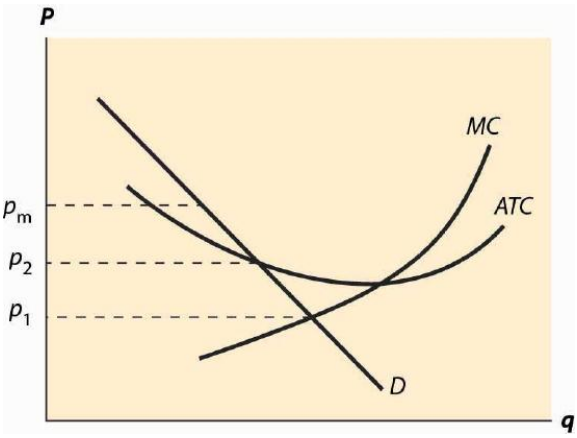


Figure 15.3 Natural monopoly

There are two strategies for limiting the price: **price-cap regulation**, which directly imposes a maximum price, and **rate of return regulation**, which limits the profitability of firms. Both of these approaches induce some inefficiency of production. In both cases, an increase in average cost may translate into additional profits for the firm, causing regulated firms to engage in unnecessary activities.

Key Takeaways

- A natural monopoly arises when a single firm can efficiently serve the entire market.
- Historically, the United States and other nations have regulated natural monopolies including electricity, telephony, and water service.
- The efficient price is typically unsustainable because of decreasing average cost.
- Efficient prices can be achieved with subsidies that have been used, for example, in postal and passenger rail services in the United States and historically for several products in Canada and Europe, including airlines and airplane manufacture. Alternatively, regulation could be imposed to limit the price to average cost, the lowest break-even price. This is the more common strategy in the United States.

- Two common strategies for limiting the price are price-cap regulation, which directly imposes a maximum price, and rate of return regulation, which limits the profitability of firms. Both of these approaches induce some inefficiency of production.

This page titled 15.6: Natural Monopoly is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

15.7: Peak-load Pricing 15.7: Peak-load PricingEdit Update titlecancel

Learning Objectives

- How do monopolies respond to predictable cost fluctuation as it arises in electricity and hotel markets?

Fluctuations in demand often require holding capacity, which is used only a fraction of the time. Hotels have off-seasons when most rooms are empty. Electric power plants are designed to handle peak demand, usually on hot summer days, with some of the capacity standing idle on other days. Demand for transatlantic airline flights is much higher in the summer than during the rest of the year. All of these examples have the similarity that an amount of capacity—hotel space, airplane seats, electricity generation—will be used over and over, which means that it is used in both high demand and low demand states. How should prices be set when demand fluctuates? This question can be reformulated as to how to allocate the cost of capacity across several time periods when demand systematically fluctuates.

Consider a firm that experiences two costs: a capacity cost and a marginal cost. How should capacity be priced? This issue applies to a wide variety of industries, including pipelines, airlines, telephone networks, construction, electricity, highways, and the Internet.

The basic **peak-load pricing** problem, pioneered by Marcel Boiteux (1922–), considers two periods. The firm's profits are given by

$$\pi = p1q1 + p2q2 - \beta max q1, q2 - mc(q1 + q2).$$

Setting price equal to marginal cost is not sustainable because a firm selling with price equal to marginal cost would not earn a return on the capacity, and thus would lose money and go out of business. Consequently, a capacity charge is necessary. The question of peak-load pricing is how the capacity charge should be allocated. This question is not trivial because some of the capacity is used in both periods.

For the sake of simplicity, we will assume that demands are independent; that is, q_1 is independent of p_2, and vice versa. This assumption is often unrealistic, and generalizing it actually doesn't complicate the problem too much. The primary complication is in computing the social welfare when demands are functions of two prices. Independence is a convenient starting point.

Social welfare is $W = \int 0q1p1(x)dx + \int 0q2p2(x)dx - \beta max q1, q2 - mc(q1 + q2).$

The Ramsey problem is to maximize W subject to a minimum profit condition. A technique for accomplishing this maximization is to instead maximize L = W + λπ.

By varying λ, we vary the importance of profits to the maximization problem, which will increase the profit level in the solution as λ increases. Thus, the correct solution to the constrained maximization problem is the outcome of the maximization of L, for some value of λ.

A useful notation is 1A, which is known as the indicator function of the set A. This is a function that is 1 when A is true, and zero otherwise. Using this notation, the first-order condition for the maximization of L is

$$0 = \partial L \partial q1 = p1(qq) - \beta 1q1 \geq q2 - mc + \lambda(p1(qq) + q1p1'(q1) - \beta 1q1 \geq q2 - mc) or p1(q1) - \beta 1q1 \geq q2 - mcp1 = \lambda\lambda + 11\varepsilon 1$$

,

where 1 q 1 ≥ q 2 is the characteristic function of the event $q_1 \geq q_2$. Similarly, $p2(q2) - \beta 1q1 \leq q2 - mcp2 = \lambda\lambda + 11\varepsilon 2$. Note as before that $\lambda \to \infty$ yields the monopoly solution.

There are two potential types of solutions. Let the demand for Good 1 exceed the demand for Good 2. Either $q_1 > q_2$, or the two are equal.

Case 1 ($q_1 > q_2$):

$$\text{p } 1(q1) - \beta - mcp1 = \lambda\lambda + 11\varepsilon 1 \text{ and } p2(q2) - mcp2 = \lambda\lambda + 11\varepsilon 2$$

In Case 1, with all of the capacity charge allocated to Good 1, quantity for Good 1 still exceeds quantity for Good 2. Thus, the peak period for Good 1 is an extreme peak. In contrast, Case 2 arises when assigning the capacity charge to Good 1 would reverse the peak—assigning all of the capacity charge to Good 1 would make Period 2 the peak.

Case 2 ($q_1 = q_2$):

The profit equation can be written $p_1(q) - mc + p_2(q) - mc = \beta$. This equation determines q, and prices are determined from demand.

The major conclusion from peak-load pricing is that either the entire cost of capacity is allocated to the peak period or there is no peak period, in the sense that the two periods have the same quantity demanded given the prices. That is, either the prices equalize the quantity demanded or the prices impose the entire cost of capacity only on one peak period.

Moreover, the price (or, more properly, the markup over marginal cost) is proportional to the inverse of the elasticity, which is known as Ramsey pricing.

Key Takeaways

- Fluctuations in demand often require holding capacity, which is used only a fraction of the time. Peak-load pricing allocates the cost of capacity across several time periods when demand systematically fluctuates.
- Important industries with peak-load problems include pipelines, airlines, telephone networks, construction, electricity, highways, and the Internet.
- Under efficient peak-load pricing, either the prices equalize the quantity demanded, or the prices impose the entire cost of capacity only on one peak period. Moreover, the markup over marginal cost is proportional to the inverse of the elasticity.

EXERCISE

1. For each of the following items, state whether you would expect peak-load pricing to equalize the quantity demanded across periods or impose the entire cost of capacity on the peak period. Explain why.
    1. Hotels in Miami
    2. Electricity

This page titled 15.7: Peak-load Pricing is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

16: Games and Strategic Behavior

Competitive theory studies price-taking consumers and firms—that is, people who can't individually affect the transaction prices. The assumption that market participants take prices as given is justified only when there are many competing participants. We have also examined monopoly, precisely because a monopoly, by definition, doesn't have to worry about competitors. Strategic behavior involves the examination of the intermediate case, where there are few enough participants that they take each other into account—and their actions individually matter—so that the behavior of any one participant influences choices of the other participants. That is, participants are strategic in their choices of action, recognizing that their choices will affect choices made by others.

The right tool for the job of examining strategic behavior in economic circumstances is game theory, the study of how people play games. Game theory was pioneered by the mathematical genius John von Neumann (1903–1957). Game theory has also been very influential in the study of military strategy; and, indeed, the strategy of the cold war between the United States and the Soviet Union was guided by game-theoretic analyses.An important reference for game theory is John von Neumann (1903–1957) and Oskar Morgenstern (1902–1977), *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press, 1944). Important extensions were introduced by John Nash (1928–), the mathematician made famous by Sylvia Nasar's delightful book, *A Beautiful Mind* (Simon & Schuster, 1998). Finally, applications in the military arena were pioneered by Nobel laureate Thomas Schelling (1921–), *The Strategy of Conflict* (Cambridge: Cambridge University Press, 1960).

The theory provides a description that fits common games like poker or the board game Monopoly, but will cover many other situations as well. In any game, there is a list of players. Games generally unfold over time; at each moment in time, players have information—possibly incomplete—about the current state of play and a set of actions they can take. Both information and actions may depend on the history of the game prior to that moment. Finally, players have payoffs and are assumed to play in such a way as to maximize their anticipated payoff, taking into account their expectations for the play of others. When the players, their information and available actions, and payoffs have been specified, we have a game.

16.1: Matrix Games
16.2: Nash Equilibrium
16.3: Mixed Strategies
16.4: Examples
16.5: Subgame Perfection
16.6: Supergames

This page titled 16: Games and Strategic Behavior is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

16.1: Matrix Games 16.1: Matrix GamesEdit Update titlecancel

Learning Objectives

- How are games modeled?
- What is optimal play?

The simplest game is called a matrix payoff game with two players. In a **matrix payoff game**, all actions are chosen simultaneously. It is conventional to describe a matrix payoff game as played by a row player and a column player. The row player chooses a row in a matrix; the column player simultaneously chooses a column. The outcome of the game is a pair of payoffs where the first entry is the payoff of the row player, and the second is the payoff of the column player. Figure 16.1.1 provides an example of a "2 × 2" matrix payoff game—the most famous game of all—which is known as the **prisoner's dilemma**. In the game, the strategies are to confess or not to confess.

		Column	
		Confess	Don't
Row	Confess	(–10, –10)	(0, –20)
	Don't	(–20, 0)	(–1, –1)

Figure 16.1.1 The prisoner's dilemma

In the prisoner's dilemma, two criminals named Row and Column have been apprehended by the police and are being questioned separately. They are jointly guilty of the crime. Each player can choose either to confess or not. If Row confesses, we are in the top row of the matrix (corresponding to the row labeled "Confess"). Similarly, if Column confesses, the payoff will be in the relevant column. In this case, if only one player confesses, that player goes free and the other serves 20 years in jail. (The entries correspond to the number of years lost to prison. The first entry is always Row's payoff; the second entry is Column's payoff.) Thus, for example, if Column confesses and Row does not, the relevant payoff is the first column and the second row.

		Column	
		Confess	Don't
Row	Confess	(–10, –10)	(0, –20)
	Don't	(–20, 0)	(–1, –1)

Figure 16.1.2 Solving the prisoner's dilemma

If Column confesses and Row does not, Row loses 20 years, and Column loses no years; that is, it goes free. This is the payoff (–20, 0) in reverse color in Figure 16.1.2. If both confess, they are both convicted and neither goes free, but they only serve 10 years each. Finally, if neither confesses, there is a 10% chance that they are convicted anyway (using evidence other than the confession), in which case they each average a year lost.

The prisoner's dilemma is famous partly because it is readily solvable. First, Row has a strict advantage to confessing, no matter what Column is going to do. If Column confesses, Row gets –10 for confessing, –20 for not confessing, and thus is better off confessing. Similarly, if Column doesn't confess, Row gets 0 for confessing (namely, goes free), –1 for not confessing, and is better off confessing. Either way, no matter what Column does, Row should choose to confess.If Row and Column are friends and care about each other, that should be included as part of the payoffs. Here, there is no honor or friendship among thieves, and Row and Column only care about what they themselves will get. This is called a **dominant strategy**, a strategy that is optimal no matter what the other players do.

The logic is exactly similar for Column: No matter what Row does, Column should choose to confess. That is, Column also has a dominant strategy to confess. To establish this, first consider what Column's best action is, when Column thinks Row will confess.

Then consider Column's best action when Column thinks Row won't confess. Either way, Column gets a higher payoff (lower number of years lost to prison) by confessing.

The presence of a dominant strategy makes the prisoner's dilemma particularly easy to solve. Both players should confess. Note that this gets them 10 years each in prison, and thus isn't a very good outcome from their perspective; but there is nothing they can do about it in the context of the game, because for each the alternative to serving 10 years is to serve 20 years. This outcome is referred to as (Confess, Confess), where the first entry is the row player's choice, and the second entry is the column player's choice.

		Piuny: Enter	Piuny: Don't
MS	Enter	(2, –2)	(5, 0)
	Don't	(0, 5)	(0, 0)

Figure 16.1.3 An entry game

Consider an entry game played by Microsoft (the row player) and Piuny (the column player), a small start-up company. Both Microsoft and Piuny are thinking about entering a new market for an online service. Figure 16.1.3 provides the payoff structure.

In this case, if both companies enter, Microsoft ultimately wins the market, earning 2 and Piuny loses 2. If either firm has the market to itself, it gets 5 and the other firm gets zero. If neither enters, they both get zero. Microsoft has a dominant strategy to enter: It gets 2 when Piuny enters, 5 when Piuny doesn't, and in both cases it does better than when it doesn't enter. In contrast, Piuny does not have a dominant strategy: Piuny wants to enter when Microsoft doesn't, and vice versa. That is, Piuny's optimal strategy depends upon Microsoft's action; or, more accurately, Piuny's optimal strategy depends upon what Piuny believes Microsoft will do.

Piuny can understand Microsoft's dominant strategy if it knows the payoffs of Microsoft.It isn't so obvious that one player will know the payoffs of another player, which often causes players to try to signal that they are going to play a certain way—that is, to demonstrate commitment to a particular advantageous strategy. Such topics are taken up in business strategy and managerial economics. Thus, Piuny can conclude that Microsoft is going to enter, and this means that Piuny should not enter. Thus, the equilibrium of the game is for Microsoft to enter and Piuny not to enter. This equilibrium is arrived at by the **iterated elimination of dominated strategies**, eliminating strategies by sequentially removing strategies that are dominated for a player. First, we eliminated Microsoft's dominated strategy in favor of its dominant strategy. Microsoft had a dominant strategy to enter, which means that the strategy of not entering was dominated by the strategy of entering, so we eliminated the dominated strategy. That leaves a simplified game in which Microsoft enters, as shown in Figure 16.1.4.

		Piuny: Enter	Piuny: Don't
MS	Enter	(2, –2)	(5, 0)

Figure 16.1.4 Eliminating a dominated strategy

In this simplified game, after the elimination of Microsoft's dominated strategy, Piuny also has a dominant strategy: not to enter. Thus, we iterate and eliminate dominated strategies again—this time eliminating Piuny's dominated strategies—and wind up with a single outcome: Microsoft enters, and Piuny doesn't. The iterated elimination of dominated strategies solves the game.A strategy may be dominated not by any particular alternate strategy but by a randomization over other strategies, which is an advanced topic not considered here.

Figure 16.1.5 shows another game, with three strategies for each player.

		Column		
		Left	Center	Right
Row	Top	(–5, –1)	(2, 2)	(3, 3)
	Middle	(1, –3)	(1, 2)	(1, 1)
	Bottom	(0, 10)	(0, 0)	(0, –10)

Figure 16.1.5 A 3 x 3 game

he process of iterated elimination of dominated strategies is illustrated in Figure 16.1.6 by actually eliminating the rows and columns, as follows. A reverse color (white text on black background) indicates a dominated strategy.

Middle dominates Bottom for Row, yielding:

		Column		
		Left	Center	Right
Row	Top	(–5, –1)	(2, 2)	(3, 3)
	Middle	(1, –3)	(1, 2)	(1, 1)
	Bottom	(0, 10)	(0, 0)	(0, –10)

Figure 16.1.6 Eliminating a dominated strategy

With Bottom eliminated, Left is now dominated for Column by either Center or Right, which eliminates the Left Column. This is shown in Figure 16.1.7.

		Column		
		Left	Center	Right
Row	Top	(–5, –1)	(2, 2)	(3, 3)
	Middle	(1, –3)	(1, 2)	(1, 1)
	Bottom	(0, 10)	(0, 0)	(0, –10)

Figure 16.1.7 Eliminating another dominated strategy

With Left and Bottom eliminated, Top now dominates Middle for Row, as shown in Figure 16.1.8.

		Column		
		Left	Center	Right
Row	Top	(–5, –1)	(2, 2)	(3, 3)
	Middle	(1, –3)	(1, 2)	(1, 1)
	Bottom	(0, 10)	(0, 0)	(0, –10)

Figure 16.1.8 Eliminating a third dominated strategy

Finally, as shown in Figure 16.1.9, Column chooses Right over Center, yielding a unique outcome after the iterated elimination of dominated strategies, which is (Top, Right).

		Column	
Row	Left	Center	Right
Top	(–5, –1)	(2, 2)	(3, 3)
Middle	(1, –3)	(1, 2)	(1, 1)
Bottom	(0, 10)	(0, 0)	(0, –10)

Figure 16.1.9 Game solved

The iterated elimination of dominated strategies is a useful concept, and when it applies, the predicted outcome is usually quite reasonable. Certainly it has the property that no player has an incentive to change his or her behavior given the behavior of others. However, there are games where it doesn't apply, and these games require the machinery of a Nash equilibrium, named for Nobel laureate John Nash (1928–).

Key Takeaways

- Strategic behavior arises where there are few enough market participants that their actions individually matter, and where the behavior of any one participant influences choices of the other participants.
- Game theory is the study of how people play games. A game consists of the players, their information and available actions, and payoffs.
- In a matrix payoff game, all actions are chosen simultaneously. The row player chooses a row in a matrix; the column player simultaneously chooses a column. The outcome of the game is a pair of payoffs where the first entry is the payoff of the row player, and the second is the payoff of the column player.
- In the prisoner's dilemma, two criminals named Row and Column have been apprehended by the police and are being questioned separately. They are jointly guilty of the crime. Each player can choose either to confess or not. Each player individually benefits from confessing, but together they are harmed.
- A dominant strategy is a strategy that is best for a player no matter what others choose.
- Iterated elimination of dominated strategies first removes strategies dominated by others, then checks if any new strategies are dominated and removes them, and so on. In many cases, iterated elimination of dominated strategies solves a game.

This page titled 16.1: Matrix Games is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

16.2: Nash Equilibrium 16.2: Nash EquilibriumEdit Update titlecancel

Learning Objectives

- What is an equilibrium to a game?

In a **Nash equilibrium**, each player chooses the strategy that maximizes his or her expected payoff, given the strategies employed by others. For matrix payoff games with two players, a Nash equilibrium requires that the row chosen maximize the row player's payoff (given the column chosen by the column player) and the column, in turn, maximize the column player's payoff (given the row selected by the row player). Let us consider first the prisoner's dilemma, which we have already seen. Here it is illustrated once again in Figure 16.2.1.

		Column	
		Confess	Don't
Row	Confess	(–10, –10)	(0, –20)
	Don't	(–20, 0)	(–1, –1)

Figure 16.2.1 Prisoner's dilemma again

Given that the row player has chosen to confess, the column player also chooses to confess because –10 is better than –20. Similarly, given that the column player chooses confession, the row player chooses confession because –10 is better than –20. Thus, for both players to confess is a Nash equilibrium. Now let us consider whether any other outcome is a Nash equilibrium. In any other outcome, at least one player is not confessing. But that player could get a higher payoff by confessing, so no other outcome could be a Nash equilibrium.

The logic of dominated strategies extends to Nash equilibrium, except possibly for ties. That is, if a strategy is strictly dominated, it can't be part of a Nash equilibrium. On the other hand, if it involves a tied value, a strategy may be dominated but still be part of a Nash equilibrium.

The Nash equilibrium is justified as a solution concept for games as follows. First, if the players are playing a Nash equilibrium, no one has an incentive to change his or her play or to rethink his or her strategy. Thus, the Nash equilibrium has a "steady state" in that no one wants to change his or her own strategy given the play of others. Second, other potential outcomes don't have that property: If an outcome is not a Nash equilibrium, then at least one player has an incentive to change what he or she is doing. Outcomes that aren't Nash equilibria involve mistakes for at least one player. Thus, sophisticated, intelligent players may be able to deduce each other's play, and play a Nash equilibrium.

Do people actually play Nash equilibria? This is a controversial topic and mostly beyond the scope of this book, but we'll consider two well-known games: tic-tac-toe (see, for example, http://www.mcafee.cc/Bin/tictactoe/index.html) and chess. Tic-tac-toe is a relatively simple game, and the equilibrium is a tie. This equilibrium arises because each player has a strategy that prevents the other player from winning, so the outcome is a tie. Young children play tic-tac-toe and eventually learn how to play equilibrium strategies, at which point the game ceases to be very interesting since it just repeats the same outcome. In contrast, it is known that chess has an equilibrium, but no one knows what it is. Thus, at this point, we don't know if the first mover (white) always wins, or if the second mover (black) always wins, or if the outcome is a draw (neither is able to win). Chess is complicated because a strategy must specify what actions to take, given the history of actions, and there are a very large number of potential histories of the game 30 or 40 moves after the start. So we can be quite confident that people are not (yet) playing Nash equilibria to the game of chess.

The second most famous game in game theory is a coordination game called **the battle of the sexes**. The battle of the sexes involves a married couple who are going to meet each other after work but haven't decided where they are meeting. Their options are a Baseball game or the Ballet. Both prefer to be with each other, but the Man prefers the Baseball game and the Woman prefers the Ballet. This gives payoffs as shown in Figure 16.2.2.

		Woman: Baseball	Woman: Ballet
Man	Baseball	(3, 2)	(1, 1)
	Ballet	(0, 0)	(2, 3)

Figure 16.2.2 The battle of the sexes

The Man would prefer that they both go to the Baseball game, and the Woman prefers that both go to the Ballet. They each get 2 payoff points for being with each other, and an additional point for being at their preferred entertainment. In this game, iterated elimination of dominated strategies eliminates nothing. One can readily verify that there are two Nash equilibria: one in which they both go to the Baseball game and one in which they both go to the Ballet. The logic is this: If the Man is going to the Baseball game, the Woman prefers the 2 points she gets at the Baseball game to the single point she would get at the Ballet. Similarly, if the Woman is going to the Baseball game, the Man gets three points going there versus zero at the Ballet. Hence, going to the Baseball game is one Nash equilibrium. It is straightforward to show that for both to go to the Ballet is also a Nash equilibrium and, finally, that neither of the other two possibilities in which they go to separate places is an equilibrium.

Now consider the game of **matching pennies**, a child's game in which the sum of the payoffs is zero. In this game, both the row player and the column player choose heads or tails, and if they match, the row player gets the coins, while if they don't match, the column player gets the coins. The payoffs are provided in Figure 16.2.3.

		Column: Heads	Column: Tails
Row	Heads	(1, –1)	(–1, 1)
	Tails	(–1, 1)	(1, –1)

Figure 16.2.3 Matching pennies

You can readily verify that none of the four possibilities represents a Nash equilibrium. Any of the four involves one player getting –1; that player can convert –1 to 1 by changing his or her strategy. Thus, whatever the hypothesized equilibrium, one player can do strictly better, contradicting the hypothesis of a Nash equilibrium. In this game, as every child who plays it knows, it pays to be unpredictable, and consequently players need to randomize. Random strategies are known as **mixed strategies** because the players mix across various actions.

Key Takeaways

- In a Nash equilibrium, each player chooses the strategy that maximizes his or her expected payoff, given the strategies employed by others. Outcomes that aren't Nash equilibria involve mistakes for at least one player.
- The game called "the battle of the sexes" has two Nash equilibria.
- In the game of matching pennies, none of the four possibilities represents a Nash equilibrium. Consequently, players need to randomize. Random strategies are known as mixed strategies because the players mix across various actions.

This page titled 16.2: Nash Equilibrium is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

16.3: Mixed Strategies 16.3: Mixed StrategiesEdit Update titlecancel

Learning Objectives

- What games require or admit randomization as part of their solution?

Let us consider the matching pennies game again, as illustrated in Figure 16.3.1.

		Column: Heads	Column: Tails
Row	Heads	(1, –1)	(–1, 1)
	Tails	(–1, 1)	(1, –1)

Figure 16.3.1 Matching pennies again

Suppose that Row believes Column plays Heads with probability p. Then if Row plays Heads, Row gets 1 with probability p and –1 with probability (1 – p), for an expected value of 2p – 1. Similarly, if Row plays Tails, Row gets –1 with probability p (when Column plays Heads), and 1 with probability (1 – p), for an expected value of 1 – 2p. This is summarized in Figure 16.3.2.

If $2p–1 > 1–2p$, then Row is better off, on average, playing Heads than Tails. Similarly, if $2p–1 < 1–2p$, then Row is better off playing Tails than Heads. If, on the other hand, $2p–1 = 1–2p$, then Row gets the same payoff no matter what Row does. In this case, Row could play Heads, could play Tails, or could flip a coin and randomize Row's play.

A **mixed strategy Nash equilibrium** involves at least one player playing a randomized strategy and no player being able to increase his or her expected payoff by playing an alternate strategy. A Nash equilibrium in which no player randomizes is called a **pure strategy Nash equilibrium**.

		Column: Heads	Column: Tails	
Row	Heads	(1, –1)	(–1, 1)	$1p + –1(1 – p) = 2p – 1$
	Tails	(–1, 1)	(1, –1)	$–1p + 1(1 – p) = 1 – 2p$

Figure 16.3.2 Mixed strategy in matching pennies

Note that randomization requires equality of expected payoffs. If a player is supposed to randomize over strategy *A* or strategy *B*, then both of these strategies must produce the same expected payoff. Otherwise, the player would prefer one of them and wouldn't play the other.

Computing a mixed strategy has one element that often appears confusing. Suppose that Row is going to randomize. Then Row's payoffs must be equal for all strategies that Row plays with positive probability. But that equality in Row's payoffs doesn't determine the probabilities with which Row plays the various rows. Instead, that equality in Row's payoffs will determine the probabilities with which Column plays the various columns. The reason is that it is Column's probabilities that determine the expected payoffs for Row; if Row is going to randomize, then Column's probabilities must be such that Row is willing to randomize.

Thus, for example, we computed the payoff to Row of playing Heads, which was $2p – 1$, where p was the probability that Column played Heads. Similarly, the payoff to Row of playing Tails was $1 – 2p$. Row is willing to randomize if these are equal, which solves for $p = ½$.

Now let's try a somewhat more challenging example and revisit the battle of the sexes. Figure 16.3.3 illustrates the payoffs once again.

		Woman: Baseball	Woman: Ballet
Man	Baseball	(3, 2)	(1, 1)
	Ballet	(0, 0)	(2, 3)

Figure 16.3.3: Mixed strategy in battle of the sexes

This game has two pure strategy Nash equilibria: (Baseball, Baseball) and (Ballet, Ballet). Is there a mixed strategy? To compute a mixed strategy, let the Woman go to the Baseball game with probability p, and the Man go to the Baseball game with probability q. Figure 16.3.4 contains the computation of the mixed strategy payoffs for each player.

		Woman: Baseball (p)	Woman: Ballet ($1-p$)	Man's E Payoff
Man	Baseball (q)	(3, 2)	(1, 1)	$3p + 1(1-p) = 1 + 2p$
	Ballet ($1-q$)	(0, 0)	(2, 3)	$0p + 2(1-p) = 2 - 2p$
	Woman's E Payoff	$2q + 0(1-q) = 2q$	$1q + 3(1-q) = 3 - 2q$	

Figure 16.3.4 Full computation of the mixed strategy

For example, if the Man (row player) goes to the Baseball game, he gets 3 when the Woman goes to the Baseball game (probability p), and otherwise gets 1, for an expected payoff of $3p + 1(1–p) = 1 + 2p$. The other calculations are similar, but you should definitely run through the logic and verify each calculation.

A mixed strategy in the battle of the sexes game requires both parties to randomize (since a pure strategy by either party prevents randomization by the other). The Man's indifference between going to the Baseball game and to the Ballet requires $1 + 2p = 2–2p$, which yields p = ¼. That is, the Man will be willing to randomize which event he attends if the Woman is going to the Ballet ¾ of the time, and otherwise to the Baseball game. This makes the Man indifferent between the two events because he prefers to be with the Woman, but he also likes to be at the Baseball game. To make up for the advantage that the game holds for him, the Woman has to be at the Ballet more often.

Similarly, in order for the Woman to randomize, the Woman must get equal payoffs from going to the Baseball game and going to the Ballet, which requires $2q = 3–2q$, or q = ¾. Thus, the probability that the Man goes to the Baseball game is ¾, and he goes to the Ballet ¼ of the time. These are independent probabilities, so to get the probability that both go to the Baseball game, we multiply the probabilities, which yields 3/16. Figure 16.3.5 fills in the probabilities for all four possible outcomes.

		Woman	
		Baseball	Ballet
Man	Baseball	3/16	9/16
	Ballet	1/16	3/16

Figure 16.3.5 Mixed strategy probabilities

Note that more than half of the time (Baseball, Ballet) is the outcome of the mixed strategy and the two people are not together. This lack of coordination is generally a feature of mixed strategy equilibria. The expected payoffs for both players are readily computed as well. The Man's payoff is $1 + 2p = 2 - 2p$, and since p = ¼, the Man obtains 1½. A similar calculation shows that the Woman's payoff is the same. Thus, both do worse than coordinating on their less preferred outcome. But this mixed strategy Nash equilibrium, undesirable as it may seem, is a Nash equilibrium in the sense that neither party can improve his or her own payoff, given the behavior of the other party.

In the battle of the sexes, the mixed strategy Nash equilibrium may seem unlikely; and we might expect the couple to coordinate more effectively. Indeed, a simple call on the telephone should rule out the mixed strategy. So let's consider another game related to the battle of the sexes, where a failure of coordination makes more sense. This is the game of "chicken." In this game, two players drive toward one another, trying to convince the other to yield and ultimately swerve into a ditch. If both swerve into the ditch, we'll call the outcome a draw and both get zero. If one swerves and the other doesn't, the driver who swerves loses and the other driver wins, and we'll give the winner one point.Note that adding a constant to a player's payoffs, or multiplying that player's payoffs by a positive constant, doesn't affect the Nash equilibria—pure or mixed. Therefore, we can always let one outcome for each player be zero, and another outcome be one. The only remaining question is what happens when neither yield, in which case a crash results. In this version, the payoff has been set at four times the loss of swerving, as shown in Figure 16.3.6, but you can change the game and see what happens.

		Column	
		Swerve	Don't
Row	Swerve	(0, 0)	(–1, 1)
	Don't	(1, –1)	(–4, –4)

Figure 16.3.6 Chicken

This game has two pure strategy equilibria: (Swerve, Don't) and (Don't, Swerve). In addition, it has a mixed strategy. Suppose that Column swerves with probability p. Then Row gets $0p + -1(1-p)$ from swerving, $1p + -4(1-p)$ from not swerving, and Row will randomize if these are equal, which requires $p = ¾$. That is, the probability that Column swerves in a mixed strategy equilibrium is ¾. You can verify that the row player has the same probability by setting the probability that Row swerves equal to q and computing Column's expected payoffs. Thus, the probability of a collision is 1/16 in the mixed strategy equilibrium.

The mixed strategy equilibrium is more likely, in some sense, in this game: If the players already knew who was going to yield, they wouldn't actually need to play the game. The whole point of the game is to find out who will yield, which means that it isn't known in advance. This means that the mixed strategy equilibrium is, in some sense, the more reasonable equilibrium.

		Column		
		Paper	Scissors	Rock
Row	Paper	(0, 0)	(–1, 1)	(1, –1)
	Scissors	(1, –1)	(0, 0)	(–1, 1)
	Rock	(–1, 1)	(1, –1)	(0, 0)

Figure 16.3.7: Rock, paper, scissors

"Rock, paper, scissors" is a child's game in which two children use their hands to simultaneously choose paper (hand held flat), scissors (hand with two fingers protruding to look like scissors), or rock (hand in a fist). The nature of the payoffs is that paper beats rock, rock beats scissors, and scissors beats paper. This game has the structure that is illustrated in Figure 16.3.7.

Key Takeaways

- A mixed strategy Nash equilibrium involves at least one player playing a randomized strategy and no player being able to increase his or her expected payoff by playing an alternate strategy.
- A Nash equilibrium without randomization is called a pure strategy Nash equilibrium.
- If a player is supposed to randomize over two strategies, then both must produce the same expected payoff.
- The matching pennies game has a mixed strategy and no pure strategy.
- The battle of the sexes game has a mixed strategy and two pure strategies.
- The game of chicken is similar to the battle of the sexes and, like it, has two pure strategies and one mixed strategy.

EXERCISES

1. Let q be the probability that Row plays Heads. Show that Column is willing to randomize, if and only if $q = ½$. (Hint: First compute Column's expected payoff when Column plays Heads, and then compute Column's expected payoff when Column plays Tails. These must be equal for Column to randomize.)
2. Show that in the rock, paper, scissors game there are no pure strategy equilibria. Show that playing all three actions with equal likelihood is a mixed strategy equilibrium.
3. Find all equilibria of the following games:

1.

Row	Column: Left	Column: Right
Up	(3, 2)	(11, 1)
Down	(4, 5)	(8, 0)

2.

Row	Column: Left	Column: Right
Up	(3, 3)	(0, 0)
Down	(4, 5)	(8, 0)

3.

Row	Column: Left	Column: Right
Up	(0, 3)	(3, 0)
Down	(4, 0)	(0, 4)

4.

Row	Column: Left	Column: Right
Up	(7, 2)	(0, 9)
Down	(8, 7)	(8, 8)

5.

Row	Column: Left	Column: Right
Up	(1, 1)	(2, 4)
Down	(4, 1)	(3, 2)

6.

Row	Column: Left	Column: Right
Up	(4, 2)	(2, 3)
Down	(3, 8)	(1, 5)

Figure 16.3.8

4. If you multiply a player's payoff by a positive constant, the equilibria of the game do not change. Is this true or false, and why?

This page titled 16.3: Mixed Strategies is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

16.4: Examples 16.4: ExamplesEdit Update titlecancel

Learning Objectives

- How can game theory be applied to the economic settings?

Our first example concerns public goods. In this game, each player can either contribute or not. For example, two roommates can either clean their apartment or not. If they both clean, the apartment is nice. If one cleans, then that roommate does all of the work and the other gets half of the benefits. Finally, if neither cleans, neither is very happy. This suggests the following payoffs as shown in Figure 16.4.1.

		Column	
		Clean	Don't
Row	Clean	(10, 10)	(0, 15)
	Don't	(15, 0)	(2, 2)

Figure 16.4.1 Cleaning the apartment

You can verify that this game is similar to the prisoner's dilemma in that the only Nash equilibrium is the pure strategy in which neither player cleans. This is a game-theoretic version of the tragedy of the commons—even though both roommates would be better off if both cleaned, neither do. As a practical matter, roommates do solve this problem, using strategies that we will investigate when we consider dynamic games.

		Column	
		Left	Right
Row	Left	(1, 1)	(0, 0)
	Right	(0, 0)	(1, 1)

Figure 16.4.2 Driving on the right

As illustrated in Figure 16.4.2 "Driving on the right", in the "driving on the right" game, the important consideration about which side of the road that cars drive on is not necessarily the right side but the same side. If both players drive on the same side, then they each get one point; otherwise, they get zero. You can readily verify that there are two pure strategy equilibria, (Left, Left) and (Right, Right), and a mixed strategy equilibrium with equal probabilities. Is the mixed strategy reasonable? With automobiles, there is little randomization. On the other hand, people walking down hallways often seem to randomize, whether they pass on the left or the right, and sometimes do that little dance where they try to get past each other—one going left and the other going right, then both simultaneously reversing, unable to get out of each other's way. That dance suggests that the mixed strategy equilibrium is not as unreasonable as it seems in the automobile application.Continental Europe drove on the left side of the road until about the time of the French Revolution. At that time, some individuals began driving on the right as a challenge to royalty, who were on the left, essentially playing the game of chicken with the ruling class. Driving on the right became a symbol of disrespect for royalty. The challengers won out, forcing a shift to driving on the right. Besides which side one drives on, another coordination game involves whether one stops or goes on red. In some locales, the tendency for a few extra cars to proceed as a traffic light changes from green to yellow to red forces those whose light changes to green to wait; and such a progression can lead to the opposite equilibrium, where one goes on red and stops on green. Under Mao Tse-tung, the Chinese considered changing the equilibrium to proceeding on red and stopping on green (because "red is the color of progress"), but wiser heads prevailed and the plan was scrapped.

		NYC	
		No Concession	Tax Rebate
LA	No Concession	(30, 10)	(10, 20)
	Tax Rebate	(20, 10)	(20, 0)

Figure 16.4.3: Bank location game

Consider a foreign bank that is looking to open a main office and a smaller office in the United States. As shown in Figure 16.4.3 "Bank location game", the bank narrows its choice for main office to either New York City (NYC) or Los Angeles (LA), and is leaning toward Los Angeles. If neither city does anything, Los Angeles will get $30 million in tax revenue and New York will get $10 million. New York, however, could offer a $10 million rebate, which would swing the main office to New York; but then New York would only get a net of $20 million. The discussions are carried on privately with the bank. Los Angeles could also offer the concession, which would bring the bank back to Los Angeles.

		Republican	
		Clean	Mud
Democrat	Clean	(3, 1)	(1, 2)
	Mud	(2, 1)	(2, 0)

Figure 16.4.4 Political mudslinging

You have probably had the experience of trying to avoid encountering someone, whom we will call Rocky. In this instance, Rocky is actually trying to find you. Here it is Saturday night and you are choosing which party, of two possible parties, to attend. You like Party 1 better and, if Rocky goes to the other party, you get 20. If Rocky attends Party 1, you are going to be uncomfortable and get 5. Similarly, Party 2 is worth 15, unless Rocky attends, in which case it is worth 0. Rocky likes Party 2 better (these different preferences may be part of the reason why you are avoiding him), but he is trying to see you. So he values Party 2 at 10, Party 1 at 5, and your presence at the party he attends is worth 10. These values are reflected in Figure 16.4.5.

		Rocky	
		Party 1	Party 2
You	Party 1	(5, 15)	(20, 10)
	Party 2	(15, 5)	(0, 20)

Figure 16.4.5 Avoiding Rocky

Our final example involves two firms competing for customers. These firms can either price High or Low. The most money is made if they both price High; but if one prices Low, it can take most of the business away from the rival. If they both price Low,

they make modest profits. This description is reflected in Figure 16.4.6.

		Firm 2	
		High	Low
Firm 1	High	(15, 15)	(0, 25)
	Low	(25, 0)	(5, 5)

Figure 16.4.6 Price cutting game

Key Takeaways

- The free-rider problem of public goods with two players can be formulated as a game.
- Whether to drive on the right or the left is a game similar to battle of the sexes.
- Many everyday situations are reasonably formulated as games.

EXERCISES

1. Verify that the bank location game has no pure strategy equilibria and that there is a mixed strategy equilibrium where each city offers a rebate with probability ½.
2. Show that the only Nash equilibrium of the political mudslinging game is a mixed strategy with equal probabilities of throwing mud and not throwing mud.
3. Suppose that voters partially forgive a candidate for throwing mud in the political mudslinging game when the rival throws mud, so that the (Mud, Mud) outcome has payoff (2.5, 0.5). How does the equilibrium change?
4. 1. Show that there are no pure strategy Nash equilibria in the avoiding Rocky game.
   2. Find the mixed strategy Nash equilibria.
   3. Show that the probability that you encounter Rocky is 7 12 .
5. Show that the firms in the price-cutting game have a dominant strategy to price low, so that the only Nash equilibrium is (Low, Low).

This page titled 16.4: Examples is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

16.5: Subgame Perfection 16.5: Subgame PerfectionEdit Update titlecancel

Learning Objectives

- How do dynamic games play out?

So far, we have considered only games that are played simultaneously. Several of these games—notably the price cutting and apartment cleaning games—are actually played over and over again. Other games, like the bank location game, may only be played once, but nevertheless are played over time. Recall the bank location game, as illustrated once again in Figure 16.5.1.

		NYC	
		No Concession	Tax Rebate
LA	No Concession	(30, 10)	(10, 20)
	Tax Rebate	(20, 10)	(20, 0)

Figure 16.5.1 Bank location game revisited

In this game, NYC makes the first move and chooses Rebate (to the left) or No Rebate (to the right). If NYC chooses Rebate, LA can then choose Rebate or None. Similarly, if NYC chooses No Rebate, LA can choose Rebate or None. The payoffs [using the standard of (LA, NYC) ordering] are written below the choices.

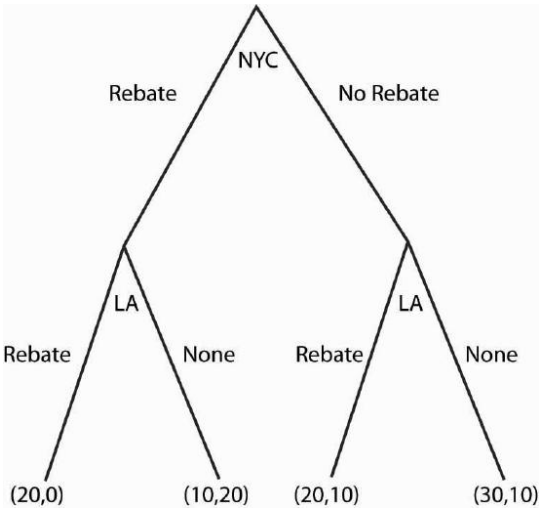


Figure 16.5.2: Sequential bank location (NYC payoff listed first)

What NYC would like to do depends upon what NYC believes LA will do. What should NYC believe about LA? (Boy, does that rhetorical question suggest a lot of facetious answers.) The natural belief is that LA will do what is in LA's best interest. This idea —that each stage of a dynamic game is played in an optimal way—is called **subgame perfection**.

Subgame perfection requires each player to act in its own best interest, independent of the history of the game.Subgame perfection was introduced by Nobel laureate Reinhard Selten (1930–). This seems very sensible and, in most contexts, it is sensible. In some settings, it may be implausible. Even if I see a player make a particular mistake three times in a row, subgame perfection requires that I must continue to believe that that player will not make the mistake again. Subgame perfection may be implausible in some circumstances, especially when it pays to be considered somewhat crazy.

In the example, subgame perfection requires LA to offer a Rebate when NYC does (since LA gets 20 by rebating vs. 10), and to not offer a Rebate when NYC doesn't. This is illustrated in the game, as shown in Figure 16.5.3, using arrows to indicate LA's choices. In addition, the actions that LA won't choose have been recolored in a light gray.

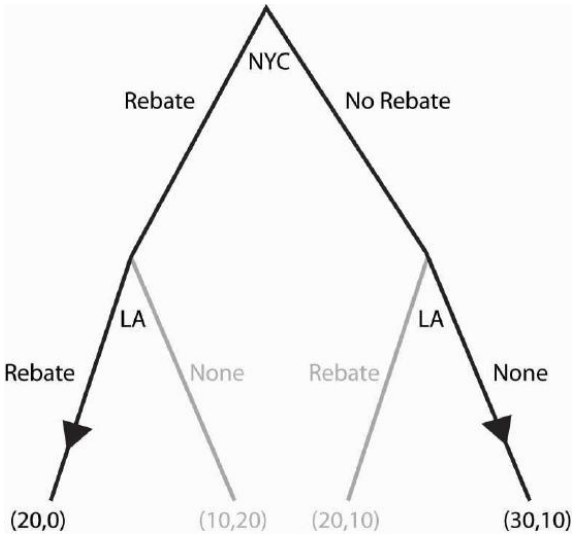


Figure 16.5.3 Subgame perfection

Once LA's subgame perfection choices are taken into account, NYC is presented with the choice of offering a Rebate, in which case it gets 0, or not offering a Rebate, in which case it gets 10. Clearly the optimal choice for NYC is to offer No Rebate, in which case LA doesn't either; and the result is 30 for LA, and 10 for NYC.

Dynamic games are generally "solved backward" in this way. That is, first establish what the last player does, then figure out—based upon the last player's expected behavior—what the penultimate player does, and so on.

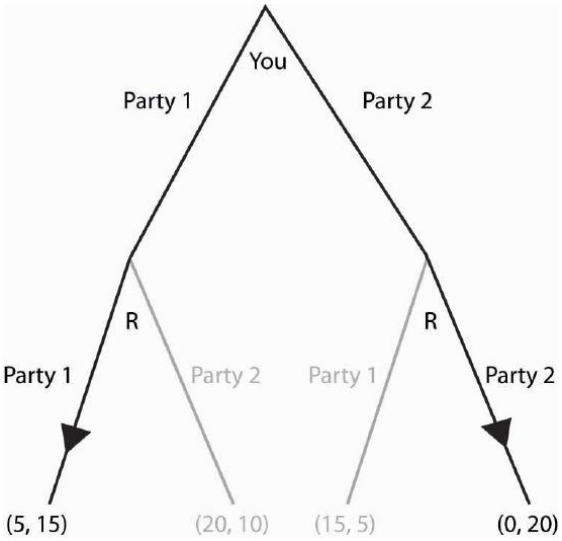


Figure 16.5.4: Can't avoid Rocky

Since Rocky's optimal choice eliminates your best outcomes, you make the best of a bad situation by choosing Party 1. Here, Rocky has a **second mover advantage**: Rocky's ability to condition on your choice means that by choosing second he does better than he would do in a simultaneous game. In contrast, a **first mover advantage** is a situation where choosing first is better than choosing simultaneously. First mover advantages arise when going first influences the second mover advantageously.

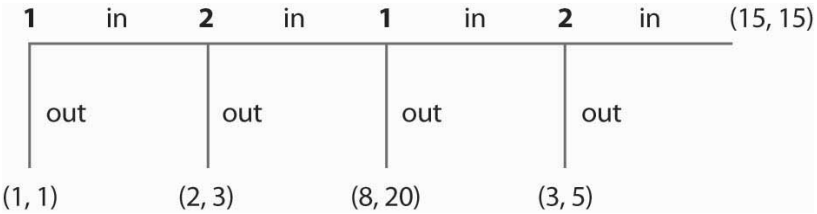
Key Takeaways

- To decide what one should do in a sequential game, one figures out what will happen in the future, and then works backward to decide what to do in the present.
- Subgame perfection requires each player to act in his or her own best interest, independent of the history of the game.
- A first mover advantage is a situation where choosing first is better than choosing simultaneously. First mover advantages arise when going first influences the second mover advantageously.
- A second mover advantage is a situation where choosing second is better than choosing simultaneously. Second mover advantages arise when going second permits exploiting choices made by others.

EXERCISES

1. Formulate the battle of the sexes as a sequential game, letting the woman choose first. (This situation could arise if the woman were able to leave a message for the man about where she has gone.) Show that there is only one subgame perfect equilibrium, that the woman enjoys a first mover advantage over the man, and that she gets her most preferred outcome.
2. What payoffs would players receive if they played this two-player sequential game below? Payoffs are listed in parentheses, with Player 1's payoffs always listed first. (Note that choosing "in" allows the other player to make a decision, while choosing "out" ends the game.)

   Figure 16.5.5



3. Consider the following game:

   Figure 16.5.6

		Column: L	Column: R
Row	U	(1, 3)	(3, 2)
	D	(4, 1)	(2, 4)

   1. Find all equilibria of the above game.
   2. What is the subgame perfect equilibrium if you turn this into a sequential game, with Column going first? With Row going first?
   3. In which game does Column get the highest payoff—the simultaneous game, the sequential game when Column goes first, or the sequential game when Column goes second?

This page titled 16.5: Subgame Perfection is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

16.6: Supergames 16.6: SupergamesEdit Update titlecancel

Learning Objectives

- What can happen in games that are repeated over and over?
- What role does the threat of retaliation play?

Some situations, like the price-cutting game or the apartment cleaning game, are played over and over. Such situations are best modeled as a **supergame**.The supergame was invented by Robert Aumann (1930–) in 1959. A supergame is a game that is played an infinite number of times, where the players discount the future. The game played each time is known as a **stage game**. Generally supergames are played in times 1, 2, 3, ….

Cooperation may be possible in supergames, if the future is important enough. Consider the price-cutting game introduced previously and illustrated again in Figure 16.6.1.

		Firm 2	
		High	Low
Firm 1	High	(15, 15)	(0, 25)
	Low	(25, 0)	(5, 5)

Figure 16.6.1: Price cutting game revisited

The dominant strategy equilibrium to this game is (Low, Low). It is clearly a subgame perfect equilibrium for the players to just play (Low, Low) over and over again because, if that is what Firm 1 thinks that Firm 2 is doing, Firm 1 does best by pricing Low, and vice versa. But that is not the only equilibrium to the supergame.

Consider the following strategy, called a **grim trigger strategy**, which involves being nice initially but not nice forever when someone else isn't cooperative. Price High, until you see your rival price Low. After your rival has priced Low, price Low forever. This is called a trigger strategy because an action of the other player (pricing Low) triggers a change in behavior. It is a grim strategy because it punishes forever.

If your rival uses a grim trigger strategy, what should you do? Basically, your only choice is when to price Low because, once you price Low, your rival will price Low, and then your best choice is also to price Low from then on. Thus, your strategy is to price High up until some point $t - 1$, and then price Low from time t on. Your rival will price High through t, and price Low from $t + 1$ on. This gives a payoff to you of 15 from period 1 through $t - 1$, 25 in period t, and then 5 in period $t + 1$ on. We can compute the payoff for a discount factor δ:

$$\begin{aligned} V_t &= 15(1+\delta+\delta^2+\ldots+\delta^{t-1})+25\delta^t+5(\delta t+1+\delta^{t+2}+\ldots) \\ &= 15\frac{1-\delta^t}{1-\delta}+25\delta^t+5\frac{\delta^t}{1-\delta} \\ &= \frac{15}{1-\delta}-\frac{\delta^t}{1-\delta}(15-25(1-\delta)-5\delta) \\ &= \frac{15}{1-\delta}-\frac{\delta^t}{1-\delta}(-10+20\delta). \end{aligned}$$

If $-10+20\delta<0$, it pays to price Low immediately, at $t = 0$, because it pays to price Low; and the earlier that one prices Low, the higher the present value. If $-10+20\delta>0$, it pays to wait forever to price Low; that is, $t = \infty$. Thus, in particular, the grim trigger strategy is an optimal strategy for a player when the rival is playing the grim trigger strategy if $\delta \geq ½$. In other words, cooperation in pricing is a subgame perfect equilibrium if the future is important enough; that is, the discount factor δ is high enough.

The logic of this example is that the promise of future cooperation is valuable when the future itself is valuable, and that promise of future cooperation can be used to induce cooperation today. Thus, Firm 1 doesn't want to cut price today because that would lead Firm 2 to cut price for the indefinite future. The grim trigger strategy punishes price cutting today with future Low profits.

Supergames offer more scope for cooperation than is illustrated in the price-cutting game. First, more complex behavior is possible. For example, consider the game shown in Figure 16.6.2:

		Firm 2	
		High	Low
Firm 1	High	(10, 10)	(0, 25)
	Low	(25, 0)	(5, 5)

Figure 16.6.2: A variation of the price-cutting game

Here, again, the unique equilibrium in the stage game is (Low, Low). But the difference between this game and the previous game is that the total profits of Firms 1 and 2 are higher in either (High, Low) or (Low, High) than in (High, High). One solution is to alternate between (High, Low) and (Low, High). Such alternation can also be supported as an equilibrium, using the grim trigger strategy—that is, if a firm does anything other than what it is supposed to do in the alternating solution, the firms instead play (Low, Low) forever.

The **folk theorem** says that if the value of the future is high enough, any outcome that is individually rational can be supported as an equilibrium to the supergame. **Individual rationality** for a player in this context means that the outcome offers a present value of profits at least as high as that offered in the worst equilibrium in the stage game from that player's perspective. Thus, in the price-cutting game, the worst equilibrium of the stage game offered each player 5, so an outcome can be supported if it offers each player at least a running average of 5.

The simple logic of the folk theorem is this. First, any infinite repetition of an equilibrium of the stage game is itself a subgame perfect equilibrium. If everyone expects this repetition of the stage game equilibrium, no one can do better than to play his or her role in the stage game equilibrium every period. Second, any other plan of action can be turned into a subgame perfect equilibrium merely by threatening any agent who deviates from that plan with an infinite repetition of the worst stage game equilibrium from that agent's perspective. That threat is credible because the repetition of the stage game equilibrium is itself a subgame perfect equilibrium. Given such a grim trigger–type threat, no one wants to deviate from the intended plan.

The folk theorem is a powerful result and shows that there are equilibria to supergames that achieve very good outcomes. The kinds of coordination failures that we saw in the battle of the sexes, and the failure to cooperate in the prisoner's dilemma, need not arise; and cooperative solutions are possible if the future is sufficiently valuable.

However, it is worth noting some assumptions that have been made in our descriptions of these games—assumptions that matter but are unlikely to be true in practice. First, the players know their own payoffs. Second, they know their rival's payoffs. They possess a complete description of the available strategies and can calculate the consequences of these strategies—not just for themselves but also for their rivals. Third, each player maximizes his or her expected payoff; they know that their rivals do the same; they know that their rivals know that everyone maximizes; and so on. The economic language for this is the structure of the game, and the players' preferences are common knowledge. Few real-world games will satisfy these assumptions exactly. Since the success of the grim trigger strategy (and other strategies we haven't discussed) generally depends upon such knowledge, informational considerations may cause cooperation to break down. Finally, the folk theorem shows us that there are lots of equilibria to supergames but provides no guidance on which ones will be played. These assumptions can be relaxed, although they may lead to wars on the equilibrium path "by accident"—and a need to recover from such wars—so that the grim trigger strategy becomes suboptimal.

Key Takeaways

- A supergame is a game that is played over and over again without end, where the players discount the future. The game played each time is known as a stage game.
- Playing a "one-shot" Nash equilibrium to the stage game forever is a subgame perfect equilibrium to the supergame.
- A grim trigger strategy involves starting play by using one behavior and, if another player ever does something else, switching to one-shot Nash behavior forever.
- The folk theorem says that if the value of the future is high enough, any outcome that is individually rational can be supported as an equilibrium to the supergame. Individual rationality for a player means that the outcome offers a present value of profits at

least as high as that offered in the worst equilibrium in the stage game from that player's perspective.
- If players are patient, full cooperation is obtainable as one of many subgame perfect equilibria to supergames.

EXERCISE

1. Consider the game in Figure 16.6.2, and consider a strategy in which Firm 1 prices High in odd-numbered periods and Low in even-numbered periods, while Firm 2 prices High in even-numbered periods and Low in odd-numbered periods. If either deviates from these strategies, both firms price Low from then on. Let δ be the discount factor. Show that these firms have a payoff of $251 - \delta 2$ or $25\delta 1 - \delta 2$, depending upon which period it is. Then show that the alternating strategy is sustainable if $10 + 5\delta 1 - \delta \leq 25\delta 1 - \delta 2$. This, in turn, is equivalent to $\delta \geq 6 - 2$.

This page titled 16.6: Supergames is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

17: Imperfect Competition

17.1: Cournot Oligopoly
17.2: Cournot Industry Performance
17.3: Hotelling Differentiation
17.4: The Circle Model

This page titled 17: Imperfect Competition is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

17.1: Cournot Oligopoly 17.1: Cournot OligopolyEdit Update titlecancel

Learning Objectives

- How do industries with only a few firms behave?
- How is their performance measured?

The Cournot oligopoly model is the most popular model of imperfect competition. It is a model in which the number of firms matters, and it represents one way of thinking about what happens when the world is neither perfectly competitive nor a monopoly.

In the **Cournot model**, there are n firms, who simultaneously set quantities. We denote a typical firm as firm i and number the firms from i = 1 to i = n. Firm i chooses a quantity $q_i \geq 0$ to sell, and this quantity costs $c_i(q_i)$. The sum of the quantities produced is denoted by Q. The price that emerges from the competition among the firms is p(Q), and this is the same price for each firm. It is probably best to think of the quantity as really representing a capacity, and competition in prices by the firms determining a market price given the market capacity.

The profit that a firm i obtains is

$$\pi_i = p(Q)q_i - c_i(q_i).$$

Each firm chooses q_i to maximize profit. (Bear in mind that Q is the sum of the firms' quantities, so that when firm i increases its output slightly, Q goes up by the same amount.) The first-order conditions give

$$0 = \frac{\partial \pi_i}{\partial q_i} = p(Q) + p'(Q)q_i - c'_i(q_i).$$

This equation holds with equality provided $q_i > 0$. A simple thing that can be done with the first-order conditions is to rewrite them to obtain the average value of the price-cost margin:

$$\frac{p(Q) - c'_i(q_i)}{p(Q)} = -\frac{p'(Q)q_i}{p(Q)} = -\frac{Qp'(Q)}{p(Q)}\frac{q_i}{Q} = \frac{s_i}{\varepsilon}.$$

Here $s_i = q_i Q$ is firm i's market share. Multiplying this equation by the market share and summing over all firms $i = 1, \ldots, n$ yields

$$\sum_{i=1}^{n} \frac{p(Q) - c'_i(q_i)}{p(Q)} s_i = \frac{1}{\varepsilon}\sum_{i=1}^{n} s_i^2 = \frac{HHI}{\varepsilon}$$

where $HHI = \sum_{i=1}^{n} s_i^2$ is the **Hirschman-Herfindahl Index (HHI)**. The HHI is named for Albert Hirschman (1915–), who invented it in 1945, and Orris Herfindahl (1918–1972), who invented it independently in 1950. The HHI has the property that if the firms are identical, so that $s_i = 1/n$ for all i, then the HHI is also 1/n. For this reason, antitrust economists will sometimes use 1/HHI as a proxy for the number of firms, and describe an industry with "2 ½ firms," meaning an HHI of 0.4.To make matters more confusing, antitrust economists tend to state the HHI using shares in percent, so that the HHI is on a 0 to 10,000 scale.

We can draw several inferences from these equations. First, larger firms, those with larger market shares, have a larger deviation from competitive behavior (price equal to marginal cost). Small firms are approximately competitive (price nearly equals marginal cost), while large firms reduce output to keep the price higher, and the amount of the reduction, in price-cost terms, is proportional to market share. Second, the HHI reflects the deviation from perfect competition on average; that is, it gives the average proportion by which price equal to marginal cost is violated. Third, the equation generalizes the "inverse elasticity result" proved for monopoly, which showed that the price-cost margin was the inverse of the elasticity of demand. The generalization states that the weighted average of the price-cost margins is the HHI over the elasticity of demand.

Because the price-cost margin reflects the deviation from competition, the HHI provides a measure of how large a deviation from competition is present in an industry. A large HHI means the industry "looks like monopoly." In contrast, a small HHI looks like perfect competition, holding constant the elasticity of demand.

The case of a symmetric (identical cost functions) industry is especially enlightening. In this case, the equation for the first-order condition can be rewritten as

$$0 = p(Q) + p'(Q)\frac{Q}{n} - c'\left(\frac{Q}{n}\right)$$

or

$$p(Q) = \frac{\varepsilon n}{\varepsilon n - 1} c'\left(\frac{Q}{n}\right).$$

Thus, in the symmetric model, competition leads to pricing as if demand was more elastic, and indeed is a substitute for elasticity as a determinant of price.

Key Takeaways

- Imperfect competition refers to the case of firms that individually have some price-setting ability or "market power" but are constrained by rivals.
- The Cournot oligopoly model is the most popular model of imperfect competition.
- In the Cournot model, firms choose quantities simultaneously and independently, and industry output determines price through demand. A Cournot equilibrium is a Nash equilibrium to the Cournot model.
- In a Cournot equilibrium, the price-cost margin of each firm is that firm's market share divided by the elasticity of demand. Hence the share-weighted average price-cost margin is the sum of market squared market shares divided by the elasticity of demand.
- The Hirschman-Herfindahl Index (HHI) is the weighted average of the price-cost margins.
- In the Cournot model, larger firms deviate more from competitive behavior than do small firms.
- The HHI measures the industry deviation from perfect competition.
- The Cournot model generalizes the "inverse elasticity result" proved for monopoly. The HHI is one with monopoly.
- A large value for HHI means the industry "looks like monopoly." In contrast, a small HHI looks like perfect competition, holding constant the elasticity of demand.
- With n identical firms, a Cournot industry behaves like a monopoly facing a demand that is n times more elastic.

This page titled 17.1: Cournot Oligopoly is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

17.2: Cournot Industry Performance 17.2: Cournot Industry PerformanceEdit Update titlecancel

Learning Objectives

- What happens to quantity-setting firms when there are fixed costs of entry?

How does the Cournot industry perform? Let us return to the more general model, which doesn't require identical cost functions. We already have one answer to this question: the average price-cost margin is the HHI divided by the elasticity of demand. Thus, if we have an estimate of the demand elasticity, we know how much the price deviates from the perfect competition benchmark.

The general Cournot industry actually has two sources of inefficiency. First, price is above marginal cost, so there is the deadweight loss associated with unexploited gains from trade. Second, there is the inefficiency associated with different marginal costs. This is inefficient because a rearrangement of production, keeping total output the same, from the firm with high marginal cost to the firm with low marginal cost, would reduce the cost of production. That is, not only is too little output produced, but what output is produced is inefficiently produced, unless the firms are identical.

To assess the productive inefficiency, we let c ′ 1 be the lowest marginal cost. The average deviation from the lowest marginal cost, then, is

$\chi = \sum i = 1nsi(c'i - c'1) = \sum i = 1nsi(p - c'1 - (p - c'i)) = p - c'1 - \sum i = 1nsi(p - c'i) = p - c'1 - p\sum i = 1nsi(p - c'i)p = p - c'1 - p\varepsilon \sum i = 1nsi2 = p - c'1 - p\varepsilon HHI.$

Thus, while a large HHI means a large deviation from price equal to marginal cost and hence a large level of monopoly power (holding constant the elasticity of demand), a large HHI also tends to indicate greater productive efficiency—that is, less output produced by high-cost producers. Intuitively, a monopoly produces efficiently, even if it has a greater reduction in total output than other industry structures.

There are a number of caveats worth mentioning in the assessment of industry performance. First, the analysis has held constant the elasticity of demand, which could easily fail to be correct in an application. Second, fixed costs have not been considered. An industry with large economies of scale, relative to demand, must have very few firms to perform efficiently, and small numbers should not necessarily indicate the market performs poorly even if price-cost margins are high. Third, it could be that entry determines the number of firms and that the firms have no long-run market power, just short-run market power. Thus, entry and fixed costs could lead the firms to have approximately zero profits, in spite of price above marginal cost.

Using Exercise 1, suppose there is a fixed cost *F* that must be paid before a firm can enter a market. The number of firms *n* should be such that firms are able to cover their fixed costs, but add one more cost and they can't. This gives us a condition determining the number of firms *n*:

$(1 - cn + 1)2 \geq F \geq (1 - cn + 2)2$

Thus, each firm's net profits are $(1 - cn + 1)2 - F \leq (1 - cn + 1)2 - (1 - cn + 2)2 = (2n + 3)(1 - c)2(n + 1)2(n + 2)2.$

Note that the monopoly profits πm are $1/4(1 - c)^2$. Thus, with free entry, net profits are less than $(2n + 3)4(n + 1)2(n + 2)2\pi m$, and industry net profits are less than $n(2n + 3)4(n + 1)2(n + 2)2\pi m$.

Table 17.1 shows the performance of the constant-cost, linear-demand Cournot industry when fixed costs are taken into account and when they aren't. With two firms, gross industry profits are 8/9 of the monopoly profits, not substantially different from monopoly. But when fixed costs sufficient to ensure that only two firms enter are considered, the industry profits are at most 39% of the monopoly profits. This percentage—39%—is large because fixed costs could be "relatively" low, so that the third firm is just deterred from entering. That still leaves the two firms with significant profits, even though the third firm can't profitably enter. As the number of firms increases, gross industry profits fall slowly toward zero. The net industry profits, on the other hand, fall dramatically and rapidly to zero. With 10 firms, the gross profits are still about a third of the monopoly level, but the net profits are only at most 5% of the monopoly level.

Table 17.1 Industry Profits as a Fraction of Monopoly Profits

Number of Firms	Gross Industry Profits (%)	Net Industry Profits (%)
2	88.9	39.0

Number of Firms	Gross Industry Profits (%)	Net Industry Profits (%)
3	75.0	27.0
4	64.0	19.6
5	55.6	14.7
10	33.1	5.3
15	23.4	2.7
20	18.1	1.6

The Cournot model gives a useful model of imperfect competition, a model that readily permits assessing the deviation from perfect competition. The Cournot model embodies two kinds of inefficiency: (a) the exercise of monopoly power and (b) technical inefficiency in production. In settings involving entry and fixed costs, care must be taken in applying the Cournot model.

Key Takeaways

- The Cournot industry has two sources of inefficiency: too little output is produced, and what output is produced is inefficiently produced (unless the firms are identical).
- The HHI analysis has held constant the elasticity of demand, which could easily fail to be correct, and fixed costs have not been considered.
- Consideration of fixed costs reduces the apparent inefficiency of Cournot industry.

EXERCISES

1. Suppose the inverse demand curve is $p(Q) = 1 - Q$, and that there are n Cournot firms, each with constant marginal cost c, selling in the market.
    1. Show that the Cournot equilibrium quantity and price are $Q = n(1 - c)n + 1$ and $p(Q) = 1 + ncn + 1$
    2. Show that each firm's gross profits are $(1 - cn + 1)2$
2. Suppose the inverse demand curve is $p(Q) = 1 - Q$, and that there are n Cournot firms, each with marginal cost c selling in the market.
    1. Find the Cournot equilibrium price and quantity.
    2. Determine the gross profits for each firm.
3. What formula from the Cournot model is used in antitrust analysis? How is it used?
4. Consider n identical Cournot firms in equilibrium.
    1. Show that the elasticity of market demand satisfies ε>1/n.
    2. Is this consistent in the case when n = 1 (monopoly)?
5. The market for Satellite Radio consists of only two firms. Suppose the market demand is given by $P = 250 - Q$, where P is the price and Q is the total quantity, so Q = Q_1 + Q_2. Each firm has total costs given by $C(Q_i) = Q_i^2 + 5Qi + 200$
    1. What is the market price predicted by the Cournot duopoly model?
    2. If the industry produces a total quantity X, what allocation of quantity (with X = Q_1 + Q_2) between the two companies minimizes total cost? (Your answer should express total cost as a function of X.)
    3. If the firms merge with the cost found in b, what is the market price?

This page titled 17.2: Cournot Industry Performance is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

17.3: Hotelling Differentiation 17.3: Hotelling DifferentiationEdit Update titlecancel

Learning Objectives

- What are the types of differentiated products and how do firms selling differentiated products behave?

Breakfast cereals range from indigestible, unprocessed whole grains to boxes that are filled almost entirely with sugar, with only the odd molecule or two of grain thrown in. Such cereals are hardly good substitutes for each other. Yet similar cereals are viewed by consumers as good substitutes, and the standard model of this kind of situation is the Hotelling model.Hotelling theory is named for Harold Hotelling (1895–1973). Hotelling was the first to use a line segment to represent both the product that is sold and the preferences of the consumers who are buying the products. In the **Hotelling model**, customers' preferences are located by points on the same line segment. The same line is used to represent products. For example, movie customers are differentiated by age, and we can represent moviegoers by their ages. Movies, too, are designed to be enjoyed by particular ages. Thus, a preteen movie is unlikely to appeal very much to a 6-year-old or to a 19-year-old, while a Disney movie appeals to a 6-year-old, but less to a 15-year-old. That is, movies have a target age, and customers have ages, and these are graphed on the same line.

Figure 17.1 Hotelling Model for Breakfast Cereals

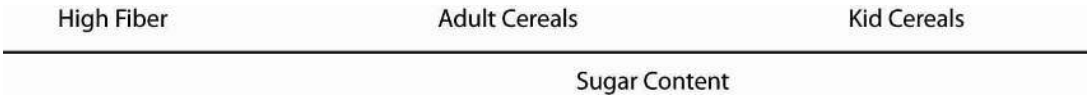


Figure 17.1. Breakfast cereals are primarily distinguished by their sugar content, which ranges on the Hotelling line from low on the left to high on the right. Similarly, the preferences of consumers also fall on the same line. Each consumer has a “most desired point,” and he or she prefers cereals closer to that point than cereals at more distant points.

There are two main types of differentiation, each of which can be modeled using the Hotelling line. These types are quality and variety. Quality refers to a situation where consumers agree on which product is better; the disagreement among consumers concerns whether higher quality is worth the cost. In automobiles, faster acceleration, better braking, higher gas mileage, more cargo space, more legroom, and greater durability are all good things. In computers, faster processing, brighter screens, higher resolution screens, lower heat, greater durability, more megabytes of RAM, and more gigabytes of hard drive space are all good things. In contrast, varieties are the elements about which there is not widespread agreement. Colors and shapes are usually varietal rather than quality differentiators. Some people like almond-colored appliances, others choose white, with blue a distant third. Food flavors are varieties, and while the quality of ingredients is a quality differentiator, the type of food is usually a varietal differentiator. Differences in music would primarily be varietal.

Quality is often called **vertical differentiation**, while variety is **horizontal differentiation**.

The standard Hotelling model fits two ice cream vendors on a beach. The vendors sell an identical product, and they can choose to locate wherever they wish. For the time being, suppose the price they charge for ice cream is fixed at $1. Potential customers are also spread randomly along the beach.

We let the beach span an interval from 0 to 1. People desiring ice cream will walk to the closest vendor because the price is the same. Thus, if one vendor locates at x and the other at y, and $x < y$, those located between 0 and ½ $(x + y)$ go to the left vendor, while the rest go to the right vendor. This is illustrated in Figure 17.2.

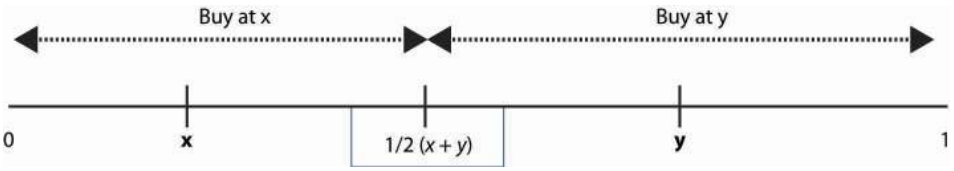


Figure 17.2 Sharing the Hotelling Market

Note that the vendor at x sells more by moving toward y, and vice versa. Such logic forces profit-maximizing vendors to both locate in the middle. The one on the left sells to everyone left of ½, while the one on the right sells to the rest. Neither can capture more of the market, so equilibrium locations have been found. (To complete the description of an equilibrium, we need to let the two “share” a point and still have one on the right side and one on the left side of that point.)

This solution is commonly used as an explanation of why U.S. political parties often seem very similar to each other—they have met in the middle in the process of chasing the most voters. Political parties can’t directly buy votes, so the “price” is fixed; the

only thing parties can do is locate their platform close to voters' preferred platform, on a scale of "left" to "right." But the same logic that a party can grab the middle, without losing the ends, by moving closer to the other party will tend to force the parties to share the same middle-of-the-road platform.

The model with constant prices is unrealistic for the study of the behavior of firms. Moreover, the two-firm model on the beach is complicated to solve and has the undesirable property that it matters significantly whether the number of firms is odd or even. As a result, we will consider a Hotelling model on a circle and let the firms choose their prices.

Key Takeaways

- In the Hotelling model, there is a line, and the preferences of each consumer are represented by a point on this line. The same line is used to represent products.
- There are two main types of differentiation: quality and variety. Quality refers to a situation where consumers agree on which product is better. Varieties are the differentiators about which there is not widespread agreement.
- Quality is often called vertical differentiation, while variety is horizontal differentiation.
- The standard Hotelling model involves two vendors selling an identical product and choosing to locate on a line. Both charge the same price. People along the line buy from the closest vendor.
- The Nash equilibrium for the standard model involves both sellers locating in the middle. This is inefficient because it doesn't minimize transport costs.
- The standard model is commonly used as a model to represent the positions of political candidates.

EXERCISE

1. Suppose there are four ice cream vendors on the beach, and customers are distributed uniformly. Show that it is a Nash equilibrium for two to locate at ¼ and two to locate at ¾.

This page titled 17.3: Hotelling Differentiation is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

17.4: The Circle Model 17.4: The Circle ModelEdit Update titlecancel

Learning Objectives

- Is there a simple, convenient model of differentiated product competition, and how does it perform?

In the **circle model**, a Hotelling model is set on a circle. There are n firms evenly spaced around the circle whose circumference is 1. Thus, the distance between any firm and each of its closest neighbors is 1/n. Consumers care about two things: how distant the firm they buy from is and how much they pay for the good. Consumers minimize the sum of the price paid and t times the distance between the consumer's location (also on the circle) and the firm. Each consumer's preference is uniformly distributed around the circle. The locations of firms are illustrated in Figure 17.3.

Figure 17.3 A Segment of the Circle Model

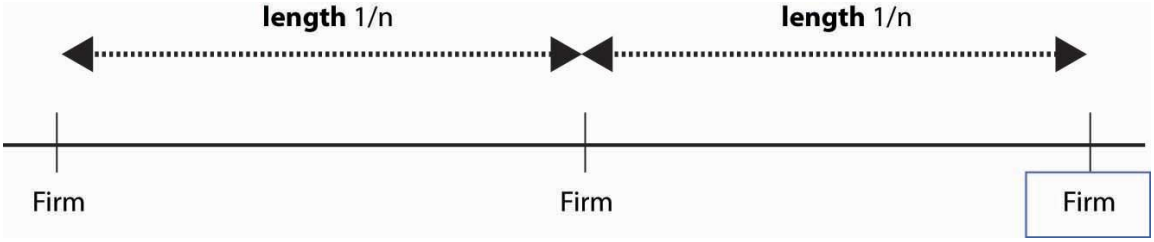


Figure 17.3 charges an alternate price r, but every other firm charges p. A consumer who is x units away from the firm pays the price r + tx from buying at the firm, or $p + t(1/n - x)$ from buying from the rival. The consumer feels indifferent toward the nearby firms if these are equal, that is, $r + tx^* = p + t\,(1/n - x^*)$ where x* is the location of the consumer who is indifferent.

$x^* = p + tn - r2t = 12n + p - r2t$

Thus, consumers who are closer than x* to the firm charging r buy from that firm, and consumers who are further away than x* buy from the alternative firm. Demand for the firm charging r is twice x* (because the firm sells to both sides), so profits are price minus marginal cost times two $x^*;\ \text{that}\ is, (r - c)2x^* = (r - c)(1n + p - rt)$

The first-order conditionBecause profit is quadratic in r, we will find a global maximum. for profit maximization is $0 = \partial\partial r(r - c)(1n + p - rt) = (1n + p - rt) - r - ct$

We could solve the first-order condition for r. But remember that the question is, when does p represent a Nash equilibrium price? The price p is an equilibrium price if the firm wants to choose r = p. Thus, we can conclude that p is a Nash equilibrium price when $p = c + tn$

This value of p ensures that a firm facing rivals who charge p also chooses to charge p. Thus, in the Hotelling model, price exceeds marginal cost by an amount equal to the value of the average distance between the firms because the average distance is 1/n and the value to a consumer for traveling that distance is t. The profit level of each firm is t n 2 , so industry profits are t n .

How many firms will enter the market? Suppose the fixed cost is F. We are going to take a slightly unusual approach and assume that the number of firms can adjust in a continuous fashion, in which case the number of firms is determined by the zero profit condition $F = tn2,\ \text{or}\ n = tF$

What is the socially efficient number of firms? The socially efficient number of firms minimizes the total costs, which are the sum of the transportation costs and the fixed costs. With n firms, the average distance a consumer travels is n $\int -12n12n|x|dx = 2n \int 012nx\,dx = n(12n)2 = 14n$

Thus, the socially efficient number of firms minimizes the transport costs plus the entry costs t 4n +nF. This occurs at n= 1 2 t F . The socially efficient number of firms is half the number of firms that enter with free entry.

Too many firms enter in the Hotelling circle model. This extra entry arises because efficient entry is determined by the cost of entry and the average distance of consumers, while prices are determined by the marginal distance of consumers, or the distance of the marginal consumer. That is, competing firms' prices are determined by the most distant customer, and that leads to prices that are too high relative to the efficient level; free entry then drives net profits to zero only when it is excess entry.

The Hotelling model is sometimes used to justify an assertion that firms will advertise too much, or engage in too much research and development (R&D), as a means of differentiating themselves and creating profits.

Key Takeaways

- A symmetric Nash equilibrium to the circle model involves a price that is marginal cost plus the transport cost t divided by the number of firms n. The profit level of each firm is t n 2 , so industry profits are t n .
- The socially efficient number of firms is half the number that would enter with free entry.
- The circle model is sometimes used to justify an assertion that firms will advertise too much, or engage in too much R&D, relative to the socially efficient amount.

This page titled 17.4: The Circle Model is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

18: Information

18.1: Market for Lemons
18.2: Myerson-Satterthwaite Theorem
18.3: Signaling
18.4: Search and Price Dispersion

This page titled 18: Information is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

18.1: Market for Lemons 18.1: Market for LemonsEdit Update titlecancel

Learning Objectives

- Can information held by sellers but relevant to buyers be an impediment to trade?

Nobel laureate George Akerlof (1940–) examined the market for used cars and considered a situation known as the **market for lemons**, where the sellers are better informed than the buyers. This is quite reasonable because sellers have owned the car for a while and are likely to know its quirks and potential problems. Akerlof showed that this differential information may cause the used car market to collapse; that is, the information possessed by sellers of used cars destroys the market and the opportunities for profitable exchange.

To understand Akerlof's insight, suppose that the quality of used cars lies on a 0 to 1 scale and that the population of used cars is uniformly distributed on the interval from 0 to 1. In addition, let that quality represent the value a seller places on the car, and suppose buyers put a value that is 50% higher than the seller. Finally, the seller knows the actual quality, while the buyer does not.

Can a buyer and seller trade in such a situation? First, note that trade is a good thing because the buyer values the car more than the seller. That is, both the buyer and seller know that they should trade. But can they agree on a price? Consider a price p. At this price, any seller who values the car less than p will be willing to trade. But because of our uniform distribution assumption, this means the distribution of qualities of cars offered for trade at price p will be uniform on the interval 0 to p. Consequently, the average quality of these cars will be ½ p, and the buyer values these cars 50% more, which yields ¾ p. Thus, the buyer is not willing to pay the price p for the average car offered at price p.

The effect of the informed seller and uninformed buyer produces a "lemons" problem. At any given price, all the lemons and only a few of the good cars are offered, and the buyer—not knowing the quality of the car—isn't willing to pay as much as the actual value of a high-value car offered for sale. This causes the market to collapse; and only the worthless cars trade at a price around zero. Economists call this situation, where some parties have information that others do not, an **informational asymmetry**.

In the real world, of course, the market has found partial or imperfect solutions to the lemons problem identified by Akerlof. First, buyers can become informed and regularly hire their own mechanic to inspect a car they are considering. Inspections reduce the informational asymmetry but are costly in their own right. Second, intermediaries offer warranties and certification to mitigate the lemons problem. The existence of both of these solutions, which involve costs in their own right, is itself evidence that the lemons problem is a real and significant problem, even though competitive markets find ways to ameliorate the problems.

An important example of the lemons problem is the inventor who creates an idea that is difficult or impossible to patent. Consider an innovation that would reduce the cost of manufacturing computers. The inventor would like to sell it to a computer company, but she or he can't tell the computer company what the innovation entails prior to price negotiations because then the computer company could just copy the innovation. Similarly, the computer company can't possibly offer a price for the innovation in advance of knowing what the innovation is. As a result, such innovations usually require the inventor to enter the computer manufacturing business, rather than selling to an existing manufacturer, entailing many otherwise unnecessary costs.

Key Takeaways

- Information itself can lead to market failures.
- The market for lemons refers to a situation where sellers are better informed than buyers about the quality of the good for sale, like used cars.
- The informational asymmetry—sellers know more than buyers—causes the market to collapse.
- Inspections, warranties, and certification mitigate the lemons problem. The existence of these costly solutions is itself evidence that the lemons problem (informational asymmetry is an impediment to trade) is a real and significant problem.
- An example of the lemons problem is the inventor who creates an idea that is difficult or impossible to patent and cannot be verified without being revealed.

EXERCISE

1. In Akerlof's market for lemons model, suppose it is possible to certify cars, verifying that they are better than a particular quality q. Thus, a market for cars "at least as good as q" is possible. What price or prices are possible in this market? (Hint: sellers offer cars only if $q \leq$ quality $\leq p$.) What quality maximizes the expected gains from trade?

This page titled 18.1: Market for Lemons is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

18.2: Myerson-Satterthwaite Theorem 18.2: Myerson-Satterthwaite TheoremEdit Update titlecancel

Learning Objectives

- Can information about values and costs that is not relevant to the other party be an impediment to trade?

The lemons problem is a situation where the buyers are relatively uninformed and care about the information held by sellers. Lemons problems are limited to situations where the buyer isn't well-informed, and these problems can be mitigated by making information public. In many transactions, the buyer knows the quality of the product, so lemons concerns aren't a significant issue. There can still be a market failure, however, if there are a limited number of buyers and sellers.

Consider the case of one buyer and one seller bargaining over the sale of a good. The buyer knows his own value v for the good, but not the seller's cost. The seller knows her own cost c for the good, but not the buyer's value. The buyer views the seller's cost as uniformly distributed on the interval [0,1], and, similarly, the seller views the buyer's value as uniformly distributed on [0,1].The remarkable fact proved by Roger Myerson and Mark Satterthwaite ("Efficient Mechanisms for Bilateral Trade," Journal of Economic Theory 28 [1983]: 265–281) is that the distributions don't matter; the failure of efficient trade is a fully general property. Philip Reny and Preston McAfee ("Correlated Information and Mechanism Design," Econometrica 60, no. 2 [March 1992]: 395–421) show the nature of the distribution of information matters, and Preston McAfee ("Efficient Allocation with Continuous Quantities," Journal of Economic Theory 53, no. 1 [February 1991]: 51–74.) showed that continuous quantities can overturn the Myerson-Satterthwaite theorem. Can efficient trade take place? Efficient trade requires that trade occurs whenever v > c, and the remarkable answer is that it is impossible to arrange efficient trade if the buyer and seller are to trade voluntarily. This is true even if a third party is used to help arrange trade, provided the third party isn't able to subsidize the transaction.

The total gains from trade under efficiency are $\int 01 \int 0vv - cdcdv = \int 01v22dv = 16.$

A means of arranging trade, known as a **mechanism**,A mechanism is a game for achieving an objective, in this case to arrange trades. asks the buyer and seller for their value and cost, respectively, and then orders trade if the value exceeds the cost and dictates a payment p by the buyer to the seller. Buyers need not make honest reports to the mechanism, however, and the mechanisms must be designed to induce the buyer and seller to report honestly to the mechanism so that efficient trades can be arranged.Inducing honesty is without loss of generality. Suppose that the buyer of type v reported the type z(v). Then we can add a stage to the mechanism in which the buyer reports a type, which is converted via the function z to a report, and then that report is given to the original mechanism. In the new mechanism, reporting v is tantamount to reporting z(v) to the original mechanism.

Consider a buyer who actually has value v but reports a value r. The buyer trades with the seller if the seller has a cost less than r, which occurs with probability r.

$$u(r,v) = vr - Ecp(r,c)$$

The buyer gets the actual value v with probability r, and makes a payment that depends on the buyer's report and the seller's report. But we can take expectations over the seller's report to eliminate it (from the buyer's perspective), and this is denoted E_c p (r, c), which is just the expected payment given the report r. For the buyer to choose to be honest, u must be maximized at r = v for every v; otherwise, some buyers would lie and some trades would not be efficiently arranged. Thus, we can concludeWe maintain an earlier notation that the subscript refers to a partial derivative, so that if we have a function f, f_1 is the partial derivative of f with respect to the first argument of $f.ddvu(v,v) = u1(v,v) + u2(v,v) = u2(v,v) = r|r = v = v.$

The value u(v,v) is the gain accruing to a buyer with value v who reports having value v. Because the buyer with value 0 gets zero, the total gain accruing to the average buyer can be computed by integrating by parts $\int 01u(v,v)dv = -(1-v)u(v,v)|v = 01 + \int 01(1 - v)(dudv)dv = \int 01(1-v)vdv = 16.$

In the integration by parts, $dv = d–(1–v)$ is used. The remarkable conclusion is that if the buyer is induced to truthfully reveal the buyer's value, the buyer must obtain the entire gains from trade. This is actually a quite general proposition. If you offer the entire gains from trade to a party, that party is induced to maximize the gains from trade. Otherwise, he or she will want to distort away from maximizing the entire gains from trade, which will result in a failure of efficiency.

The logic with respect to the seller is analogous: the only way to get the seller to report her cost honestly is to offer her the entire gains from trade.

The **Myerson-Satterthwaite theorem** shows that private information about value may prevent efficient trade. Thus, the gains from trade are insufficient to induce honesty by both parties. (Indeed, they are half the necessary amount.) Thus, any mechanism for arranging trades between the buyer and the seller must suffer some inefficiency. Generally this occurs because buyers act like they value the good less than they do, and sellers act like their costs are higher than they truly are.

It turns out that the worst-case scenario is a single buyer and a single seller. As markets get "thick," the per capita losses converge to zero, and markets become efficient. Thus, informational problems of this kind are a small-numbers issue. However, many markets do in fact have small numbers of buyers or sellers. In such markets, it seems likely that informational problems will be an impediment to efficient trade.

Key Takeaways

- The Myerson-Satterthwaite theorem shows that the gains from trade are insufficient to induce honesty about values and costs by a buyer and seller. Any mechanism for arranging trades between the buyer and the seller must suffer some inefficiency.
- Generally this inefficiency occurs because buyers act like they value the good less than they do, and sellers act like their costs are higher than they truly are, resulting in an inefficiently low level of trade.
- As markets get "thick," the per capita losses converge to zero, and markets become efficient. Informational problems of this kind are a small-numbers issue.

EXERCISE

1. Let h(r, c) be the gains of a seller who has cost c and reports r, $h(r,c) = p(v,r)–(1–r)c.$

   Noting that the highest cost seller (c = 1) never sells and thus obtains zero profits, show that honesty by the seller implies the expected value of h is 1/16.

This page titled 18.2: Myerson-Satterthwaite Theorem is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

18.3: Signaling 18.3: SignalingEdit Update titlecancel

Learning Objectives

- Why do people spend so much money on things that aren't much better than the cheap versions?
- Why do people who study ancient Greek get good jobs that don't involve reading ancient Greek?

An interesting approach to solving informational problems involves **signaling**.Signaling was introduced by Nobel laureate Michael Spence in his dissertation, part of which was reprinted in "Job Market Signaling," Quarterly Journal of Economics 87 (August 1973): 355–74. Signaling, in economic jargon, means expenditures of time or money whose purpose is to convince others of something. Thus, people signal wealth by wearing Rolex watches, driving expensive cars, or sailing in the America's Cup. They signal erudition by tossing quotes from Kafka or Tacitus into conversations. They signal being chic by wearing the "right" clothes and listening to cool music. Signaling is also rampant in the animal world, from peacock feathers to elk battles, and it is the subject of a vibrant and related research program.

A university education serves not just to educate but also to signal the ability to learn. Businesses often desire employees who are able to adapt to changing circumstances and who can easily and readily learn new strategies and approaches. Education signals such abilities because it will be easier for quick learners to perform well at university. A simple model suffices to illustrate the point. Suppose there are two types of people. Type A has a low cost cA of learning, and type B has a higher cost cB of learning. It is difficult to determine from an interview whether someone is type A or type B. Type A is worth more to businesses, and the competitive wage wA (expressed as a present value of lifetime earnings) for type A's is higher than the wage wB for type B's.

A person can signal that she is a type A by pursuing a sufficient amount of education. Suppose the person devotes an amount of time x to learning at university, thus incurring the cost cA x. If x is large enough so that $wA–cAx > wB > wA–cBx$, it pays the type A to obtain the education, but not the type B, if education in fact signals that the student is type A. Thus, a level of education x in this case signals a trait (ease of learning) that is valued by business, and it does so by voluntary choice—those with a high cost of learning choose not to obtain the education, even though they could do it. This works as a signal because only type A would voluntarily obtain the education in return for being perceived to be a type A.

There are several interesting aspects to this kind of signaling. First, the education embodied in x need not be valuable in itself; the student could be studying astronomy or ancient Greek, neither of which are very useful in most businesses but are nevertheless strong signals of the ability to learn. Second, the best subject matter for signaling is that in which the difference in cost between the type desired by employers and the less desirable type is greatest—that is, where cB – cA is greatest. Practical knowledge is somewhat unlikely to make this difference great; instead, challenging and abstract problem solving may be a better separator. Clearly, it is desirable to have the subject matter be useful, if it can still do the signaling job. But interpreting long medieval poems could more readily signal the kind of flexible mind desired in management than studying accounting, not because the desirable type is good at it or that it is useful, but because the less desirable type is so much worse at it.

Third, one interprets signals by asking, "What kinds of people would make this choice?" while understanding that the person makes the choice hoping to send the signal. Successful law firms have very fine offices, generally much finer than the offices of their clients. Moreover, there are back rooms at most law firms, where much of the real work is done, that aren't nearly so opulent. The purpose of the expensive offices is to signal success, essentially proclaiming, "We couldn't afford to waste money on such expensive offices if we weren't very successful. Thus, you should believe we are successful."

The law firm example is similar to the education example. Here, the cost of the expenditures on fancy offices is different for different law firms because more successful firms earn more money and thus value the marginal dollar less. Consequently, more successful firms have a lower cost of a given level of office luxury. What is interesting about signaling is that it is potentially quite wasteful. A student spends 4 years studying boring poems and dead languages in order to demonstrate a love of learning, and a law firm pays $75,000 for a conference table that it rarely uses and gets no pleasure out of in order to convince a client that the firm is extremely successful. In both cases, it seems like a less costly solution should be available. The student can take standardized tests, and the law firm could show its win-loss record to the potential client. But standardized tests may measure test-taking skills rather than learning ability, especially if what matters is the learning ability over a long time horizon. Win-loss records can be "massaged," and in the majority of all legal disputes, the case settles and both sides consider themselves the winner. Consequently, statistics may not be a good indicator of success, and the expensive conference table may be a better guide.

Key Takeaways

- Signaling means expenditures of time or money whose purpose is to convince others of something.
- A university education signals the ability to learn. The education need not be useful in itself to be helpful in signaling. The best subject matter for signaling is something that is easy for fast learners and difficult for slower learners.
- Signals are interpreted by asking, “What kinds of people would make this choice?” while understanding that the person makes the choice hoping to send the signal.

This page titled 18.3: Signaling is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

18.4: Search and Price Dispersion 18.4: Search and Price DispersionEdit Update titlecancel

Learning Objectives

- Why do we see such random prices at the grocery store?
- Why are prices unpredictable?

Decades ago, economists used to make a big deal about the **law of one price**, which states that identical goods sell at the same price. The argument in favor of the law of one price is theoretical. Well-informed consumers will buy identical goods from the lowest-price seller. Consequently, the only seller to make any sales is the low-price seller. This kind of consumer behavior forces all sellers to sell at the same price.

There are few markets where the law of one price is actually observed to hold. Organized exchanges, like stock, bond, and commodity markets, will satisfy the law of one price. In addition, gas stations across the street from each other will often offer identical prices, but *often* is not *always*.

Many economists believed that the Internet would force prices of standardized goods—DVD players, digital cameras, MP3 players—to a uniform, and uniformly low, price. However, this has not occurred. Moreover, it probably can't occur in the sense that pure price competition would put the firms out of business and hence can't represent equilibrium behavior.

There are many markets where prices appear unpredictable to consumers. The price of airline tickets is notorious for unpredictability. The price of milk, soft drinks, paper towels, and canned tuna varies 50% or more depending on whether or not the store has an advertised sale of the item. Prices of goods sold on the Internet vary substantially from day to day.It is often very challenging to assess Internet prices because of variations in shipping charges. Such variation of price across stores is known as **price dispersion** by economists. It is different from price discrimination because price dispersion entails a given store quoting the same price to all customers; the variation is across stores, while price discrimination varies across customers.

Why are prices so unpredictable? We now develop a model that shows they have to be.

To understand price dispersion, we divide consumers into two types: shoppers and loyal customers. Loyal customers won't pay more than a price *pm* for the good, but they consult a particular store. If that store has the good for less than the price *pm*, the loyal customer buys; otherwise, he or she does not. In contrast, the shoppers buy only from the store offering the lowest price; shoppers are informed about the prices offered by all stores. We let the proportion of shoppers be *s*. The loyal customers are allocated to the other stores equally so that if there are *n* stores, each store gets a fraction $(1 - s)/n$ of the customers. Let the marginal cost of the good be *c*, and assume that $c < pm$. Both kinds of customers buy only one unit.

For the purposes of this analysis, we will assume that prices can be chosen from the continuum. This makes the analysis more straightforward, but there is an alternate version of the analysis (not developed here) that makes the more reasonable assumption of prices that are an integer number of pennies.

First note that there is no pure strategy equilibrium. To see this, consider the lowest price *p* charged by any firm. If that price is *c*, the firm makes no money, so it would do better by raising its price to *pm* and selling only to the loyal customers. Thus, the lowest price *p* exceeds *c*. If there is a tie at *p*, it pays to break the tie by charging a billionth of a cent less than *p*, and thereby capturing all the shoppers rather than sharing them with the other firm charging *p*. So there can't be a tie.

But no tie at *p* means the next-lowest firm is charging something strictly greater than *p*, which means the lowest-price firm can increase price somewhat and not suffer any loss of sales. This contradicts profit maximization for that firm. The conclusion is that firms must randomize and that no pure strategy equilibrium exists.

But how do they randomize? We are going to look for a distribution of prices. Each firm will choose a price from the continuous distribution *F*, where $F(x)$ is the probability the firm charges a price less than *x*. What must *F* look like? We use the logic of mixed strategies: the firm must get the same profits for all prices that might actually be charged under the mixed strategy; otherwise, it would not be willing to randomize.

A firm that charges price $p \leq pm$ always sells to its captive customers. In addition, it sells to the shoppers if the other firms have higher prices, which occurs with probability $(1 - F(p))n - 1$. Thus, the firm's profits are $\pi(p) = (p - c)(1 - sn + s(1 - F(p))n - 1)$.

On each sale, the firm earns $p - c$. The firm always sells to its loyal customers and in addition captures the shoppers if the other firms price higher. Because no firm will exceed *pm*, the profits must be the same as the level arising from charging *pm*, and this gives $\pi(p) = (p-c)(1-sn+s(1-F(p))n-1) = (pm-c)1-sn.$

This equation is readily solved for *F*:

$$F(p) = (1-(pm-p)(1-s)s(p-c)n)1n-1$$

The lower bound of prices arises at the point *L* where *F*(*L*) = 0, or

$$L = c+(pm-c)1-sn1-sn+s$$

These two equations provide a continuous distribution of prices charged by each firm, which is an equilibrium to the pricing game. That is, each firm randomizes over the interval [*L*, *pm*], according to the continuous distribution *F*. Any price in the interval [*L*,*pm*] produces the same profits for each firm, so the firms are willing to randomize over this interval.

The loyal customers get a price chosen randomly from *F*, so we immediately see that the shoppers make life better for the loyal customers by pushing average price down. (An increase in *s* directly increases *F*, which means prices fall—recall that *F* gives the probability that prices are below a given level, so an increase in *F* is an increase in the probability of low prices.)

Similarly loyal customers make life worse for shoppers, increasing prices on average to shoppers. The distribution of prices facing shoppers is actually the distribution of the minimum price. Because all firms charge a price exceeding *p* with probability $(1–F(p))n$, at least one charges a price less than *p* with probability $1-(1-F(p))^n$, and this is the distribution of prices facing shoppers. That is, the distribution of prices charged to shoppers is

$1-(1-\mathrm{F(p))n} = 1-((\mathrm{pm-p})(1-\mathrm{s})\mathrm{s}(\mathrm{p-c})\mathrm{n})\mathrm{nn}-1$

How does a price-dispersed industry perform? First, average industry profits are $n\pi(p) = (pm-c)(1-s).$

An interesting aspect of this equation is that it doesn't depend on the number of firms, only on the number of loyal customers. Essentially, the industry profits are the same that it would earn as if the shoppers paid marginal cost and the loyal customers paid the monopoly price, although that isn't what happens in the industry, except in the limit as the number of firms goes to infinity. Note that this formula for industry profits does not work for a monopoly. To capture monopoly, one must set *s* = 0 because shoppers have no alternative under monopoly.

As the number of firms gets large, the price charged by any one firm converges to the monopoly price *pm*. However, the lowest price offered by any firm actually converges to *c*, marginal cost. Thus, in the limit as the number of firms gets large, shoppers obtain price equal to marginal cost and loyal firms pay the monopoly price.

Figure 18.1 Expected Prices in Search Equilibrium



Figure 18.1, letting the number of firms vary. Thus, with many firms, most of the gains created by the shoppers flow to shoppers. In contrast, with few firms, a significant fraction of the gains created by shoppers goes instead to the loyal customers.

Similarly, we can examine the average prices for loyal customers and shoppers when the proportion of shoppers varies. Increasing the proportion of shoppers has two effects. First, it makes low prices more attractive, thus encouraging price competition because capturing the shoppers is more valuable. Second, it lowers industry profits because the set of loyal customers is reduced. Figure 18.2 plots the average price for loyal customers and shoppers, as the proportion of shoppers ranges from zero to one, when there are five firms, $pm = 1$ and $c = 0$.

People who are price-sensitive and shop around convey a positive externality on other buyers by encouraging price competition. Similarly, people who are less price-sensitive and don't shop around convey a negative externality on the other buyers. In markets with dispersed information about the best prices, where some buyers are informed and some are not, randomized prices are a natural outcome. That is, randomization of prices, and the failure of the law of one price, is just a reflection of the different willingness or ability to search on the part of consumers.



Figure 18.2 Expected Prices (s = Proportion of Shoppers)

This difference in the willingness to search could arise simply because search is itself costly. That is, the shoppers could be determined by their choice to shop in such a way that the cost of shopping just balances the expected gains from searching. The proportion of shoppers may adjust endogenously to ensure that the gains from searching exactly equal the costs of searching. In this

way, a cost of shopping is translated into a randomized price equilibrium in which there is a benefit from shopping and all consumers get the same total cost of purchase on average.

Key Takeaways

- The law of one price, which states that identical goods sell at the same price, is unfortunately empirically false.
- There are many markets where prices appear unpredictable to consumers. Price variation over time or across stores is known as price dispersion.
- The basic price-dispersion model involves firms setting prices simultaneously and selling to two types of customers: one loyal to a particular store, the other ("shoppers") buying from the cheapest store.
- There is no pure strategy equilibrium because firms either want to be just barely the cheapest or at the monopoly price. The only candidate for an equilibrium is a mixed strategy.
- There is a mixed strategy involving randomizing over an interval of prices.
- Industry profits in price dispersion arise from the number of loyal customers and are independent of the number of firms.
- An increase in the number of firms is good for shoppers and bad for loyal customers. In the limit as the number of firms goes to infinity, the shoppers pay marginal cost and loyal customers pay the monopoly price.
- Shoppers convey a positive externality on other buyers by encouraging price competition.

This page titled 18.4: Search and Price Dispersion is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

19: Agency Theory

19.1: Principals and Agents
19.2: Cost of Providing Incentives
19.3: Multi-tasking without Homogeneity
19.4: Multi-tasking without Homogeneity

This page titled 19: Agency Theory is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

19.1: Principals and Agents 19.1: Principals and AgentsEdit Update titlecancel

Learning Objectives

- What is an agency?
- How should a principal compensate an agent?

An **agent** is a person who works for, or on behalf of, another. Thus, an employee is an agent of a company. But agency extends beyond employee relationships. Independent contractors are also agents. Advertising firms, lawyers, and accountants are agents of their clients. The CEO of a company is an agent of the board of directors of the company. A grocery store is an agent of the manufacturer of corn chips sold in the store. Thus, the agency relationship extends beyond the employee into many different economic relationships. The entity—person or corporation—on whose behalf an agent works is called a **principal**.

Agency theory is the study of incentives provided to agents. Incentives are an issue because agents need not have the same interests and goals as the principal. Employees spend billions of hours every year browsing the Web, e-mailing friends, and playing computer games while they are supposedly working. Attorneys hired to defend a corporation in a lawsuit have an incentive not to settle, to keep the billing flowing. (Such behavior would violate the attorneys' ethics requirements.) Automobile repair shops have been known to use substandard or used replacement parts and bill for new, high-quality parts. These are all examples of a conflict in the incentives of the agent and the goals of the principal.

Agency theory focuses on the cost of providing incentives. When you rent a car, an agency relationship is created. Even though a car rental company is called an agency, it is most useful to look at the renter as the agent because it is the renter's behavior that is an issue. The company would like the agent to treat the car as if it were her own car. The renter, in contrast, knows it isn't her own car and often drives accordingly.

rented P. J. O'Rourke, Republican Party Reptile (Boston: Atlantic Monthly, 1987), 242.

How can the car rental company ensure that you don't put its car into reverse while going forward at a high rate of speed? It could monitor your behavior, perhaps by putting a company representative in the car with you. That would be a very expensive and unpleasant solution to the problem of incentives. Instead, the company uses outcomes—if damage is done, the driver has to pay for it. That is also an imperfect solution because some drivers who abuse the cars get off scot-free, and others who don't abuse the car still have cars that break down and are then mired in paperwork while they try to prove their good behavior. That is, a rule that penalizes drivers based on outcomes imposes risk on the drivers. Modern technology is improving monitoring with GPS tracking.

To model the cost of provision of incentives, we consider an agent like a door-to-door encyclopedia salesperson. The agent will visit houses and sell encyclopedias to some proportion of the households; the more work the agent does, the more sales that are made. We let x represent the average dollar value of sales for a given level of effort; x is a choice the agent makes. However, x will come with risk to the agent, which we model using the variance δ^2.

The firm will pay the agent a share s of the money generated by the sales. In addition, the firm will pay the agent a salary y, which is fixed independently of sales. This scheme—a combination of salary and commission—covers many different situations. Real estate agents receive a mix of salary and commission. Authors receive an advance and a royalty, which works like a salary and commission.

The monetary compensation of the agent is $sx + y$. In addition, the agent has a cost of effort, which we take to be x 2 2a . Here, a represents the ability of the agent: more able agents, who have a higher value of a, have a lower cost of effort. Finally, there is a cost of risk. The actual risk imposed on the agent is proportional to the degree he shares in the proceeds. If s is small, the agent faces almost no monetary risk, but if s is high, most of the risk is imposed on the agent. We use the linear cost of risk model, developed earlier, to impose a cost of risk, which is s $\lambda\delta^2$. Here, δ^2 is the variance of the monetary risk, λ defines the agent's attitude or cost of risk, and s is the share of the risk imposed on the agent. This results in a payoff to the agent of $u = sx + y - x22a - s\lambda\sigma2$.

The part of the equation represented by sx + y is the payments made to the agent. The next term is the cost of generating that level of x. The final term is the cost of risk imposed on the agent by the contract.

The agency game works as follows. First, the principal offers a contract, which involves a commission s and a salary y. The agent can either accept or reject the contract and accepts if he obtains at least u_0 units of utility, the value of his next best offer. Then the agent decides how much effort to expend; that is, the agent chooses x.

As with all subgame perfect equilibria, we work backward to first figure out what x an agent would choose. Because our assumptions make u quadratic in x, this is a straightforward exercise, and we conclude x = sa. This can be embedded into u, and the agent's optimized utility u* is

$$u^* = s2a + y - (sa)22a - s\lambda\sigma 2 = y + 1/2s2a - s\lambda\sigma 2$$

The agent won't accept employment unless u* ≥ u_0, the reservation utility. The principal can minimize the cost of employing the agent by setting the salary such that u* = u_0, which results in

$$y = u0 - 1/2s2a + s\lambda\sigma 2$$

.

Observe that the higher the salary, the greater is the risk δ^2. That is, the principal has to cover the cost of risk in the salary term.

Key Takeaways

- An agent is a person who works for, or on behalf of, another.
- An employee is an agent of a company. Independent contractors are also agents. The entity—person or corporation—on whose behalf an agent works is called a principal.
- Agency theory is the study of incentives provided to agents. In the basic agency model, a principal sets a salary and commission, and the agent chooses the effort to expend. The principal keeps the random output minus the salary and commission. Higher commissions increase the agent's incentive but impose risk on the agent.

This page titled 19.1: Principals and Agents is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

19.2: Cost of Providing Incentives 19.2: Cost of Providing IncentivesEdit Update titlecancel

Learning Objectives

- How much does it cost to motivate agents?

The principal obtains profits, which are the remainder of the value after paying the agent minus the salary:

$$\pi = (1 - s)x - y = (1 - s)sa - (u0 - ½s2a + s\lambda\sigma2) = sa - u0 - ½s2a - s\lambda\sigma2$$

Note that the principal gets the entire output x = sa minus all the costs—the reservation utility of the agent u_0, the cost of providing effort, and the risk cost on the agent. That is, the principal obtains the full gains from trade—the value of production minus the total cost of production. However, the fact that the principal obtains the full gains from trade doesn't mean the principal induces the agent to work extremely hard because there is no mechanism for the principal to induce the agent to work hard without imposing more risk on the agent, and this risk is costly to the principal. Agents are induced to work hard by tying their pay to their performance, and such a link necessarily imposes risk on the agent, and risk is costly.There is a technical requirement that the principal's return π must be positive; otherwise, the principal would rather not contract at all. This amounts to an assumption that u_0 is not too large. Moreover, if s comes out less than zero, the model falls apart, and in this case, the actual solution is s = 0.

We take the principal to be risk neutral. This is reasonable when the principal is economically large relative to the agent, so that the risks faced by the agent are small compared to those faced by the principal. For example, the risks associated with any one car are small to a car rental company. The principal who maximizes expected profits chooses s to maximize π, which yields $s = 1 - \lambda a\sigma2$.

This formula is interesting for several reasons. First, if the agent is neutral to risk, which means λ = 0, then s is 1. That is, the agent gets 100% of the marginal return to effort, and the principal just collects a lump sum. This is reminiscent of some tenancy contracts used by landlords and peasants; the peasant paid a lump sum for the right to farm the land and then kept all of the crops grown. Because these peasants were unlikely to be risk neutral, while the landlord was relatively neutral to risk, such a contract was unlikely to be optimal. The contract with s = 1 is known as selling the agency because the principal sells the agency to the agent for a lump sum payment. (Here, y will generally be negative—the principal gets a payment rather than paying a salary.) The more common contract, however, had the landowner and the tenant farmer share the proceeds of farming, which gives rise to the name sharecropper.

Second, more risk or more risk aversion on the part of the agent decreases the share of the proceeds accruing to the agent. Thus, when the cost of risk or the amount of risk is high, the best contract imposes less risk on the agent. Total output sa falls as the costs of risk rise.

Third, more able agents (higher a) get higher commissions. That is, the principal imposes more risk on the more able agent because the returns to imposition of risk—in the form of higher output—are greater and thus worth the cost in terms of added risk.

Most real estate agencies operate on a mix of salary and commission, with commissions paid to agents averaging about 50%. The agency RE/MAX, however, pays commissions close to 100%, collecting a fixed monthly fee that covers agency expenses from the agents. RE/MAX claims that their formula is appropriate for better agents. The theory developed suggests that more able agents should obtain higher commissions. But in addition, RE/MAX's formula also tends to attract more able agents because able agents earn a higher wage under the high commission formula. (There is a potential downside to the RE/MAX formula: it discourages agency-wide cooperation.)

Consider what contracts attract what kinds of agents. For a fixed salary y and commission s, the agent's utility, optimizing over x, is $u* = y + ½s2a - s\lambda\sigma2$.

The agent's utility is increasing in a and decreasing in λ. Thus, more able agents get higher utility, and less risk-averse agents get higher utility.

How do the terms of the contract affect the pool of applicants? Let us suppose that two contracts are offered, one with a salary y_1 and commission s_1, the other with salary y_2 and commission s_2. We suppose $y_2 < y_1$ and $s_2 > s_1$. What kind of agent prefers Contract 2, the high-commission, low-salary contract, over Contract 1?

$$y2 + ½s22a - s2\lambda\sigma2 \geq y1 + ½s12a - s1\lambda\sigma2$$

or the equivalent:

$$½a(s22 - s12) - (s2 - s1)\lambda\sigma 2 \geq y1 - y2$$

Thus, agents with high ability a or low level of risk aversion λ prefer the high-commission, low-salary contract. A company that puts more of the compensation in the form of commission tends to attract more able agents and agents less averse to risk. The former is a desirable feature of the incentive scheme because more able agents produce more. The latter, the attraction of less risk-averse agents, may or may not be desirable but is probably neutral overall.

One important consideration is that agents who overestimate their ability will react the same as people who have higher ability. Thus, the contract equally attracts those with high ability and those who overestimate their ability.

Agency theory provides a characterization of the cost of providing incentives. The source of the cost is the link between incentives and risk. Incentives link pay and performance; when performance is subject to random fluctuations, linking pay and performance also links pay and the random fluctuations. Thus, the provision of incentives necessarily imposes risk on the agent, and if the agent is risk averse, this is costly.

In addition, the extent to which pay is linked to performance tends to affect the type of agent who is willing to work for the principal. Thus, a principal must not only consider the incentive to work hard created by the commission and salary structure but also the type of agent who would choose to accept such a contract.

Key Takeaways

- The principal chooses the salary to minimize the cost of the agent; thus, the principal nets the total output, minus the cost of the agent.
- The agent's cost must be at least as large as what the agent would get in an alternative occupation and thus includes a risk adjustment.
- The optimal commission offered by the principal is decreasing in the risk aversion of the agent and the level of risk and increasing in the agent's ability.
- If the agent is neutral to risk, the principal gets a lump sum, and "sells the agency."
- Total output falls as the costs of risk rise.
- A company that puts more of the compensation in the form of commission tends to attract more able agents and agents less averse to risk. A principal must not only consider the incentive to work hard created by the commission and salary structure but also the type of agent who would choose to accept such a contract.

EXERCISE

1. Describe how a principal would go about hiring agents who are willing to take risks.

This page titled 19.2: Cost of Providing Incentives is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

19.3: Multi-tasking without Homogeneity 19.3: Multi-tasking without HomogeneityEdit Update titlecancel

Learning Objectives

- How does an employer compensate an agent who performs several tasks of varying importance? What happens when the tasks conflict?

Multi-tasking refers to performing several activities simultaneously. All of us multitask. We study while drinking a caffeinated beverage; we think about things in the shower; we talk all too much on cell phones and eat french fries while driving. In the context of employees, an individual employee is assigned a variety of tasks and responsibilities, and the employee must divide her time and efforts among the tasks. Incentives provided to the employee must direct not only the total efforts of the employee, but also the allocation of time and effort across activities. An important aspect of multitasking is the interaction of incentives provided to an employee, and the effects of changes in one incentive on the behavior of the employee over many different dimensions. In this section, we will establish conditions under which the problem of an employer disaggregates; that is, the incentives for performing each individual task can be set independently of the incentives applied to the others.

This section is relatively challenging and involves a number of pieces. To simplify the presentation, some of the analyses are set aside as claims.

To begin the analysis, we consider a person who has n tasks or jobs. For convenience, we will index these activities with the natural numbers 1, 2, …, n. The level of activity, which may also be thought of as an action, in task i will be denoted by x_i. It will prove convenient to denote the vector of actions by $x = (x1, \dots, xn)$. We suppose the agent bears a cost c **(x)** of undertaking the vector of actions **x**. We make four assumptions on c:

1. c is increasing in each x_i.
2. c has a continuous second derivative.
3. c is strictly convex.
4. c is homogeneousHomogeneous functions were defined in Chapter 10. of degree r.

For example, if there are two tasks (n = 2), then all four of these assumptions are met by the cost function $c(x1, x2) = x12 + x22 + 1/2 \times 1 \times 2$. This function is increasing in x_1 and x_2, has continuous derivatives, is strictly convex (more about this below), and is homogeneous of degree 2.

It is assumed that c is increasing to identify the activities as costly. Continuity of derivatives is used for convenience. Convexity of c will ensure that a solution to the first-order conditions is actually an optimum for the employee. Formally, a function is a **convex function** such that, for any vectors **x** ≠ **y** and scalar α between zero and one

$$(0 \leq a \leq 1), ac(x) + (1 - a)c(y) \geq c(ax + (1 - a)y)$$

In other words, a convex function is any function that lies below the straight line segment connecting two points on the function, for any two points in the interval, when **x** is a scalor.

One way of interpreting this requirement is that it is less costly to do the average of two things than the average of the costs of the things. Intuitively, convexity requires that doing a medium thing is less costly than the average of two extremes. This is plausible when extremes tend to be very costly. It also means the set of vectors that cost less than a fixed amount, {**x** | c**(x)** ≤ b}, is a convex set. Thus, if two points cost less than a given budget, the line segment connecting them does, too. Convexity of the cost function ensures that the agent's optimization problem is concave and thus that the first-order conditions describe a maximum. When the inequality is strict for α satisfying 0 < α < 1, we refer to convexity as strict convexity.

The assumption of homogeneity dictates that scale works in a particularly simple manner. Scaling up activities increases costs at a fixed rate r. Homogeneity has very strong implications that are probably unreasonable in many settings. Nevertheless, homogeneity leads to an elegant and useful theory, as we shall see. Recall the definition of a homogeneous function: c is homogeneous of degree r means that for any $\lambda > 0, c(\lambda x) = \lambda r c(x)$.

Claim: Strict convexity implies that $r > 1$.

Proof of Claim: Fix any **x** and consider the two points **x** and λ **x**. By convexity, for

$$0 < a < 1, (a + (1 - a)\lambda r)c(x) = ac(x) + (1 - a)c(\lambda x) > c(ax + (1 - a)\lambda x)) = (a + (1 - a)\lambda)rc(x)$$

which implies $(a + (1 - a)\lambda r) > (a + (1 - a)\lambda)r$.

Define a function k that is the left-hand side minus the right-hand side:

$k(a) = a + (1 - a)\lambda r - (a + (1 - a)\lambda)r$. Note that $k(0) = k(1) = 0$. Moreover, $k''(a) = -r(r - 1)(a + (1 - a)\lambda)r - 2(1 - \lambda)2$. It is readily checked that if a convex function of one variable is twice differentiable, then the second derivative is greater than zero. If

$r \leq 1, k''(a) \geq 0$, implying that k is convex, and hence, if $0 < a < 1, k(a) = k((1 - a)0 + a1) \leq (1 - a)k(0) + ak(1) = 0$.

Similarly, if r > 1, k is concave and k(α) > 0. This completes the proof, showing that r ≤ 1 is not compatible with the strict convexity of c.

How should our person behave? Consider linear incentives, which are also known as piece rates. With piece rates, the employee gets a payment p_i for each unit of x_i produced. The person then chooses **x** to maximize $u = \sum i = 1npixi - c(x) = p \cdot x - c(x)$.

Here • is the dot product, which is the sum of the products of the components.

The agent chooses **x** to maximize u, resulting in n first-order conditions $\partial u \partial xi = pi - \partial c(x) \partial xi = pi - ci(x)$, where c_i is the partial derivative of c with respect to the ith argument x_i. This first-order condition can be expressed more compactly as $0 = p - c'(x)$ where c ′ (x) is the vector of partial derivatives of c. Convexity of c ensures that any solution to this problem is a global utility maximum because the function u is concave, and strict convexity ensures that there is at most one solution to the first-order conditions.This description is slightly inadequate because we haven't considered boundary conditions. Often a requirement like x_i ≥ 0 is also needed. In this case, the first-order conditions may not hold with equality for those choices where x_i = 0 is optimal.

One very useful implication of homogeneity is that incentives scale. Homogeneity has the effect of turning a very complicated optimization problem into a problem that is readily solved, thanks to this very scaling.

Claim: If all incentives rise by a scalar factor α, then **x** rises by α 1 r−1 .

Proof of Claim: Note that differentiating $c(\lambda x) = \lambda rc(x)$ with respect to x_i yields $c'(\lambda x) = \lambda r - 1c'(x)$, and thus $c'(\lambda x) = \lambda r - 1c'(x)$. That is, if c is homogeneous of degree r, c ′ is homogeneous of degree r – 1. Consequently, if $0 = p - c'(x), 0 = ap - c'(a1r - 1x)$. Thus, if the incentives are scaled up by α, the efforts rise by the scalar factor α 1 r−1 .

Now consider an employer with an agent engaging in n activities. The employer values the ith activity at v_i and thus wishes to maximize $\pi = \sum i = 1n(vi - pi)xi = \sum i = 1n(vi - ci(x))xi$.

This equation embodies a standard trick in agency theory. Think of the principal (employer) not as choosing the incentives **p**, but instead as choosing the effort levels **x**, with the incentives as a constraint. That is, the principal can be thought of as choosing **x** and then choosing the **p** that implements this **x**. The principal's expected profit is readily differentiated with respect to each x_j, yielding $0 = vj - cj(x) - \sum i = 1ncij(x))xi$.

However, because c_j**(x)** is homogeneous of degree r – 1,

$$\sum i = 1ncij(x))xi = dd\lambda cj(\lambda x)|\lambda = 1 = dd\lambda\lambda r - 1cj(x)|\lambda = 1 = (r - 1)cj(x),$$

and thus

$0 = vj - cj(x) - \Sigma i = 1nc$ ij $(x)) \times i = vj - rcj(x)$

This expression proves the main result of this section. Under the maintained hypotheses (convexity and homogeneity), an employer of a multitasking agent uses incentives that are a constant proportion of value; that is, $pj = vjr$, where r is the degree of homogeneity of the agent's costs. Recalling that r > 1, the principal uses a **sharing rule**, sharing a fixed proportion of value with the agent.

When agents have a homogeneous cost function, the principal has a very simple optimal incentive scheme, requiring quite limited knowledge of the agent's cost function (just the degree of homogeneity). Moreover, the incentive scheme works through a somewhat surprising mechanism. Note that if the value of one activity, for example, Activity 1, rises, p_1 rises and all the other payment rates stay constant. The agent responds by increasing x_1, but the other activities may rise or fall depending on how complementary they are to Activity 1. Overall, the agent's substitution across activities given the new incentive level on Activity 1 implements the desired effort levels on other activities. The remarkable implication of homogeneity is that, although the principal desires different effort levels for all activities, only the incentive on Activity 1 must change.

Key Takeaways

- Multi-tasking refers to performing several activities simultaneously.
- In the agency context, multitasking refers to the incentives of a principal to compensate different tasks.
- A simple model of multitasking provides a convex cost of a set of tasks that is homogeneous of degree r in the tasks. This means that scaling up activities increases costs at a fixed rate r.
- With piece rates, the employee gets a fixed payment for each unit produced.
- One very useful implication of homogeneity is that incentives scale. If all incentives rise by a scalar factor α, then **x** rises by α 1 r−1 , where r is the degree of homogeneity.
- Given convexity and homogeneity, an employer of a multitasking agent uses incentives that are a constant proportion of value; that is, $pj = vjr$.

This page titled 19.3: Multi-tasking without Homogeneity is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

19.4: Multi-tasking without Homogeneity 19.4: Multi-tasking without HomogeneityEdit Update titlecancel

Learning Objectives

- When will the incentives for performing tasks be related to each other, and how are they related?

In the previous section we saw, for example, that if the agent has quadratic costs, the principal pays the agent half the value of each activity. Moreover, the more rapidly costs rise in scale, the lower are the payments to the agent.

This remarkable theorem has several limitations. The requirement of homogeneity is itself an important limitation, although this assumption is reasonable in some settings. More serious is the assumption that all of the incentives are set optimally for the employer. Suppose, instead, that one of the incentives is set too high, at least from the employer's perspective. This might arise if, for example, the agent acquired all the benefits of one of the activities. An increase in the power of one incentive will then tend to spill over to the other actions, increasing for complements and decreasing for substitutes. When the efforts are substitutes, an increase in the power of one incentive causes others to optimally rise, to compensate for the reduced supply of efforts of that type.Multi-tasking (and agency theory more generally) is a rich theory with many implications not discussed here. For a challenging and important analysis, see Bengt Holmstrom and Paul Milgrom, "The Firm as an Incentive System," American Economic Review 84, no. 4 (September 1994): 972–991.

We can illustrate the effects of cost functions that aren't homogeneous in a relatively straightforward way. Suppose the cost depends on the sum of the squared activity levels:

$$c(x) = g(\sum i = 1nxi2) = g(x \bullet x).$$

This is a situation where vector notation (dot-products) dramatically simplifies the expressions. You may find it useful to work through the notation on a separate sheet, or in the margin, using summation notation to verify each step. At the moment, we won't be concerned with the exact specification of g, but instead we will use the first-order conditions to characterize the solution.

The agent maximizes

$$u = p \cdot x - g(x \cdot x)$$

This gives a first-order condition

$$0 = p - 2q'(x \cdot x)x$$

It turns out that a sufficient condition for this equation to characterize the agent's utility maximization is that g is both increasing and convex (increasing second derivative).

This is a particularly simple expression because the vector of efforts, **x**, points in the same direction as the incentive payments **p**. The scalar that gives the overall effort levels, however, is not necessarily a constant, as occurs with homogeneous cost functions. Indeed, we can readily see that **x** • **x** is the solution to p•p= (2 g ′ (x•x)) 2 (x•x).

Because **x** • **x** is a number, it is worth introducing notation for it: S = **x** • **x**. Then S is the solution to $p \cdot p = 4S\,(g'(S))\,2$.

The principal or employer chooses **p** to maximize $\pi = v \bullet x - p \bullet x = v \bullet x - 2g'(x \bullet x)(x \bullet x)$.

This gives the first-order condition $0 = v - 4(g'(x \bullet x) + (x \bullet x)g''(x \bullet x))x$.

Thus, the principal's choice of **p** is such that **x** is proportional to **v**, with constant of proportionality $g'(x \cdot x) + x \cdot xg''(x \cdot x)$. Using the same trick (dotting each side of the first-order condition $v = 4(g'(x \bullet x) + x \bullet xg''(x \bullet x))x$ with itself), we obtain

$$v \cdot v = 16\,(g'\,(S^*) + S^*g''\,(S^*))\,2S^*$$

which gives the level of $x \cdot x = S^*$ induced by the principal. Given S*, **p** is given by $-p = 2g'(x \cdot x)x = 2g'\,(S^*) \vee 4\,(g'\,(S^*) + S^*g''\,(S^*)) = 12\,(11 + S^*g''\,(S^*)\,g'\,(S^*))\,v$

Note that this expression gives the right answer when costs are homogeneous. In this case, g (S) must be in the form $S^{r/2}$, and the formula gives

$p = 12(11 + r - 1)v = vr$ as we already established.

The natural assumption to impose on the function g is that $(g'(S) + Sg''(S))\,2S$ is an increasing function of S. This assumption implies that as the value of effort rises, the total effort also rises.

Suppose S g ″ (S) g ′ (S) is increasing in S. Then an increase in v_i increases S, decreasing p_j for j ≠ i. That is, when one item becomes more valuable, the incentives for performing the others are reduced. Moreover, because $p \cdot p = 4S\,(g'(S))\,2$, an increase in S only occurs if **p** • **p** increases.

These equations together imply that an increase in any one v_i increases the total effort (as measured by $S^* = \mathrm{x} \cdot \mathrm{x}\backslash]$), increases the total incentives as measured by **p** • **p**, and decreases the incentives for performing all activities other than activity i. In contrast, if $\mathrm{Sg}''(S)\mathrm{g}'(\mathrm{S})$ is a decreasing function of S, then an increase in any one v_i causes all the incentives to rise. Intuitively, the increase in v_i directly causes p_i to rise because x_i is more valuable. This causes the agent to substitute toward activity i. This causes the relative cost of total activity to fall (because $\mathrm{Sg}''(\mathrm{S})\mathrm{g}'(\mathrm{S})$ decreases), which induces a desire to increase the other activity levels. This is accomplished by an increase in the incentives for performing the other activities.

This conclusion generalizes readily and powerfully. Suppose that $c(x) = g(h(x))$, where h is homogeneous of degree r and g is increasing. In the case just considered, $h(x) = x \cdot x$. Then the same conclusion, that the sign of d p i d v j is determined by the derivative of $\mathrm{Sg}''(S)\mathrm{g}'(\mathrm{S})$, holds. In the generalization, S now stands for h(**x**).

KEY TAKEAWAY

- In general, incentives can be substitutes or complements; that is, an increase in the importance of one activity may increase or decrease the incentives provided for performing the other activities. Homogeneity is the condition that causes such interactions to be zero.

This page titled 19.4: Multi-tasking without Homogeneity is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

20: Auctions

20.1: English Auction
20.2: Sealed-bid Auction
20.3: Dutch Auction
20.4: Vickrey Auction
20.5: The Winner’s Curse and Linkage
20.6: Auction Design

This page titled 20: Auctions is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

20.1: English Auction 20.1: English AuctionEdit Update titlecancel

Learning Objectives

- What is the most common auction form?
- How should I bid in an auction if I know my own value?
- When we share a value that none of us know, should I bid my estimate of value?

An **English auction** is the common auction form used for selling antiques, art, used cars, and cattle. The auctioneer starts low and calls out prices until no bidder is willing to bid higher than the current high price. The most common procedure is for a low price to be called out and a bidder to accept it. A higher price is called out, and a different bidder accepts it. When several accept simultaneously, the auctioneer accepts the first one spotted. This process continues until a price is called out that no one accepts. At that point, the auction ends, and the highest bidder wins.

Information plays a significant role in bidding in auctions. The two major extremes in information, which lead to distinct theories, are **private values**, which means bidders know their own value, and **common values**, in which bidders don't know their own value but have some indication or signal about the value. In the private values situation, a bidder may be outbid by another bidder but doesn't learn anything from another bidder's willingness to pay. The case of private values arises when the good being sold has a quality apparent to all bidders, no hidden attributes, and no possibility of resale. In contrast, the case of common values arises when bidders don't know the value of the item for sale, but that value is common to all. The quintessential example is an offshore oil lease. No one knows for sure how much oil can be extracted from an offshore lease, and companies have estimates of the amount of oil. The estimates are closely guarded because rivals could learn from them. Similarly, when antiques dealers bid on an antique, the value they place on it is primarily the resale value. Knowing rivals' estimates of the resale value could influence the value each bidder placed on the item.

The private values environment is unrealistic in most settings because items for sale usually have some element of common value. However, some situations approximate the private values environment and these are the most easy to understand.

In a private values setting, a very simple bidding strategy is optimal for bidders: a bidder should keep bidding until the price exceeds the value a bidder places on it, at which point the bidder should stop. That is, bidders should drop out of the bidding when the price exceeds their value because at that point, winning the auction requires the bidder to take a loss. Every bidder should be willing to continue to bid to prevent someone else from winning the auction provided the price is less than the bidder's value. If you have a value v and another bidder is about to win at a price $p_a < v$, you might as well accept a price p_b between the two, $p_a < p_b < v$ because a purchase at this price would provide you with a profit. This strategy is a dominant strategy for each private values bidder because no matter what strategy the other bidders adopt, bidding up to value is the strategy that maximizes the profit for each bidder.

The presence of a dominant strategy makes it easy to bid in the private values environment. In addition, it simplifies the analysis of the English auction relatively simple.

Most auctioneers use a flexible system of **bid increments**. A bid increment is the difference between successive price requests. The theory is simplest when the bid increment, denoted by δ, is very small. In this case, the bidder with the highest value wins, and the price is no more than the second-highest value, but it is at least the second-highest value minus δ, because a lower price would induce the bidder with the second-highest value to submit a slightly higher bid. If we denote the second-highest value with the somewhat obscure (but standard) notation $v_{(2)}$, the final price p satisfies $v(2) - \Delta \leq p \leq v(2)$.

As the bid increment gets small, the price is nailed down. The conclusion is that, when bid increments are small and bidders have private values, the bidder with the highest value wins the bidding at a price equal to the second-highest value. The notation for the highest value is $v_{(1)}$, and thus the seller obtains $v_{(2)}$, and the winning bidder obtains profits of $V_{(1)} - V_{(2)}$.

Key Takeaways

- An auction is a trading mechanism where the highest bidder wins an object. Auctions are typically used when values are uncertain, and thus information is an important aspect of analyzing auctions.
- Private values mean bidders know their own value.
- Common values mean bidders share a common but unknown value, and they have some indication or signal about the value. With common values, willingness to pay by one bidder is informative for other bidders.

- In an English auction, the auctioneer starts low and calls out prices until no bidder is willing to bid higher than the current high price. At that point the auction ends, and the highest bidder wins.
- In a private values setting, the English auction has a dominant strategy: remain bidding until one's value is reached.
- When bid increments are small and bidders have private values, the bidder with the highest value wins the bidding at a price equal to the second-highest value.

This page titled 20.1: English Auction is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

20.2: Sealed-bid Auction 20.2: Sealed-bid AuctionEdit Update titlecancel

Learning Objectives

- How should I bid if I don't get to see the bids of others?

In a **sealed-bid auction**, each bidder submits a bid in an envelope. These are opened simultaneously, and the highest bidder wins the item and pays his or her bid. Sealed-bid auctions are used to sell offshore oil leases, and they are used by governments to purchase a wide variety of items. In a purchase situation, known often as a tender, the lowest bidder wins the amount he bids.

The analysis of the sealed-bid auction is more challenging because the bidders don't have a dominant strategy. Indeed, the best bid depends on what the other bidders are bidding. The bidder with the highest value would like to bid a penny more than the next highest bidder's bid, whatever that might be.

To pursue an analysis of the sealed-bid auction, we are going to make a variety of simplifying assumptions. These assumptions aren't necessary to the analysis, but we make them to simplify the mathematical presentation.

We suppose there are *n* bidders, and we label the bidders 1, …, *n*. Bidder *i* has a private value v_i, which is a draw from the uniform distribution on the interval [0,1]. That is, if $0 \le a \le b \le 1$, the probability that bidder *i*'s value is in the interval $[a, b]$ is $b - a$. An important attribute of this assumption is symmetry—the bidders all have the same distribution. In addition, the formulation has assumed independence—the value one bidder places on the object for sale is statistically independent from the value placed by others. Each bidder knows his own value but he doesn't know the other bidders' values. Each bidder is assumed to bid in such a way as to maximize his expected profit (we will look for a Nash equilibrium of the bidding game). Bidders are permitted to submit any bid equal to or greater than zero.

To find an equilibrium, it is helpful to restrict attention to linear strategies, in which a bidder bids a proportion of her value. Thus, we suppose that each bidder bids λ*v* when her value is *v* and λ is a positive constant, usually between zero and one. With this set up we shall examine under what conditions these strategies comprise a Nash equilibrium. An equilibrium exists when all other bidders bid λ*v* when their value is *v*, and the remaining bidders bid the same.

So fix a bidder and suppose that bidder's value is v_i. What bid should the bidder choose? A bid of *b* wins the bidding if all other bidders bid less than *b*. Because the other bidders, by hypothesis, bid λ*v* when their value is *v*, our bidder wins when b≥λ v j for each other bidder *j*. This occurs when b λ ≥ v j for each other bidder *j*, and this in turn occurs with probability b λ . If b>λ, then in fact the probability is 1. You can show that no bidder would ever bid more than λ. Thus, our bidder with value v_i who bids *b* wins with probability (b λ) n−1 because the bidder must beat all *n* −1 other bidders. That creates expected profits for the bidder of $n = (vi - b)(b\lambda)n - 1$.

The bidder chooses *b* to maximize expected profits. The first-order condition requires $0 = -(b\lambda)n - 1 + (vi - b)(n - 1)bn - 2\lambda n - 1$.

The first-order condition solves for $b = n - 1nv$.

But this is a linear rule. Thus, if $\lambda = n - 1n$, we have a Nash equilibrium.

The nature of this equilibrium is that each bidder bids a fraction λ= n−1 n of his value, and the highest-value bidder wins at a price equal to that fraction of her value.

In some cases, the sealed-bid auction produces regret. Regret means that a bidder wishes she had bid differently. Recall our notation for values: $v_{(1)}$ is the highest value and $v_{(2)}$ is the second-highest value. Because the price in a sealed-bid auction is $n - 1n \vee (1)$, the second-highest bidder will regret her bid when $v(2) > n - 1n \vee (1)$. In this case, the bidder with the second-highest value could have bid higher and won, if the bidder had known the winning bidder's bid. In contrast, the English auction is regret-free: the price rises to the point that the bidder with the second-highest value won't pay.

How do the two auctions compare in prices? It turns out that statistical independence of private values implies **revenue equivalence**, which means the two auctions produce the same prices on average. Given the highest value $v_{(1)}$, the second-highest value has distribution $(v(2) \vee (1))n - 1$ because this is the probability that all *n* − 1 other bidders have values less than $v_{(2)}$. But this gives an expected value of $v_{(2)}$ of $\mathrm{E}v(2) = \int 0 \vee (1) \vee (2)(n - 1)v(2)n - 2v(1)n - 1$ d $v(2) = n - 1$ n $v(1)$.

Thus, the average price paid in the sealed-bid auction is the same as the average price in the English auction.

Key Takeaways

- In a sealed-bid auction, bids are opened simultaneously, and the highest bidder wins the item and pays his bid.
- The analysis of the sealed-bid auction is more challenging because the bidders don’t have a dominant strategy.
- When bidders have uniformly and independently distributed values, there is an equilibrium where they bid a constant fraction of value, n−1 n where *n* is the number of bidders.
- Statistical independence of private values implies revenue equivalence, which means English and sealed-bid auctions produce the same prices on average.

This page titled 20.2: Sealed-bid Auction is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

20.3: Dutch Auction 20.3: Dutch AuctionEdit Update titlecancel

Learning Objectives

- Don’t the Dutch use a different kind of auction to sell tulips?
- How does the Dutch auction work?

The **Dutch auction** is like an English auction, except that prices start high and are successively dropped until a bidder accepts the going price, and the auction ends. The Dutch auction is so-named because it is used to sell cut flowers in Holland, in the enormous flower auctions.

A strategy in a Dutch auction is a price at which the bidder bids. Each bidder watches the price decline, until it reaches such a point that either the bidder bids or a rival bids, and the auction ends. Note that a bidder could revise his bid in the course of the auction, but there isn’t any reason to do so. For example, suppose the price starts at $1,000, and a bidder decides to bid when the price reaches $400. Once the price gets to $450, the bidder could decide to revise and wait until $350. However, no new information has become available and there is no reason to revise. In order for the price to reach the original planned bid of $400, it had to reach $450, meaning that no one bid prior to a price of $450. In order for a bid of $400 to win, the price had to reach $450; if the price reaching $450 means that a bid of $350 is optimal, then the original bid of $400 could not have been optimal.Of course, a bidder who thinks losing is likely may wait for a lower price to formulate the bid, a consideration ignored here. In addition, because the Dutch auction unfolds over time, bidders who discount the future will bid slightly higher in a Dutch auction as a way of speeding it along, another small effect that is ignored for simplicity.

What is interesting about the Dutch auction is that it has exactly the same possible strategies and outcomes as the sealed-bid auction. In both cases, a strategy for a bidder is a bid, no bidder sees the others’ bids until after her own bid is formulated, and the winning bidder is the one with the highest bid. This is called **strategic equivalence**. Both games—the Dutch auction and the sealed-bid auction—offer identical strategies to the bidders and, given the strategies chosen by all bidders, produce the same payoff. Such games should produce the same outcomes.

The strategic equivalence of the Dutch auction and the sealed-bid auction is a very general result that doesn’t depend on the nature of the values of the bidders (private vs. common) or the distribution of information (independent vs. correlated). Indeed, the prediction that the two games should produce the same outcome doesn’t even depend on risk aversion, although that is more challenging to demonstrate.

Key Takeaways

- The Dutch auction is like an English auction, except that prices start high and are successively dropped until a bidder accepts the going price, at which point the auction ends.
- The Dutch auction is so-named because it is used to sell cut flowers in Holland.
- The Dutch auction has exactly the same possible strategies and outcomes as the sealed-bid auction. This is called strategic equivalence. As a result, the Dutch and sealed-bid auctions have the same equilibria.

This page titled 20.3: Dutch Auction is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

20.4: Vickrey Auction 20.4: Vickrey AuctionEdit Update titlecancel

Learning Objectives

- How should I bid in the auction used by eBay, assuming I don't want to "buy it now."

The strategic equivalence of the Dutch and sealed-bid auction suggests another fact: there may be more than one way of implementing a given kind of auction. Such logic led Nobel laureate William Vickrey (1914–1996) to design what has become known as the **Vickrey auction**, which is a second-price, sealed-bid auction. This auction is most familiar because it is the foundation of eBay's auction design. The Vickrey auction is a sealed-bid auction, but with a twist: the high bidder wins but pays the second-highest bid. This is why the Vickrey auction is called a second-price auction: the price is not the highest bid, but the second-highest bid.

The Vickrey auction underlies the eBay outcome because when a bidder submits a bid in the eBay auction, the current "going" price is not the highest bid, but the second-highest bid plus a bid increment. Thus, up to the granularity of the bid increment, the basic eBay auction is a Vickrey auction run over time.

As in the English auction, bidders with private values in a Vickrey auction have a dominant strategy. Fix a bidder, with value v, and let p be the highest bid of the other bidders. If the bidder bids b, the bidder earns profits of { 0 if b<p v−p if b>p }.

It is profitable for the bidder to win if $v > p$ and to lose if $v < p$. To win when $v > p$ and to lose if $v < p$ can be assured by bidding v. Essentially, there is no gain to bidding less than your value because your bid doesn't affect the price, only the likelihood of winning. Bidding less than value causes the bidder to lose when the highest rival bid falls between the bid and the value, which is a circumstance that the bidder would like to win. Similarly, bidding more than value creates a chance of winning only when the price is higher than the bidder's value, in which case the bidder would prefer to lose.

Thus, bidders in a Vickrey auction have a dominant strategy to bid their value. This produces the same outcome as in the English auction, however, because the payment made is the second-highest value, which was the price in the English auction. Thus, the Vickrey auction is a sealed-bid implementation of the English auction when bidders have private values, producing the same outcome, which is that the highest-value bidder wins but pays the second-highest value.

Because the Vickrey auction induces bidders to bid their value, it is said to be demand revealing. Unlike the English auction, in which the bidding stops when the price reaches the second-highest value and thus doesn't reveal the highest value, the Vickrey auction reveals the highest value. In a controlled, laboratory setting, demand revelation is useful, especially when the goal is to identify buyer values. Despite its theoretical niceties, the Vickrey auction can be politically disastrous. Indeed, New Zealand sold radio spectrum with the Vickrey auction on the basis of advice by a naïve economist, and the Vickrey auction created a political nightmare when a nationwide cellular license received a high bid of $110 million and a second-highest bid of $11 million. The political problem was that the demand revelation showed that the government received only about 10% of the value of the license, making the public quite irate. The situation dominated news coverage at the time.The Vickrey auction generally produces higher prices than regular sealed-bid auctions if bidders are symmetric (that is, share the same distribution of values), but it is a poor choice of auction format when bidders are not symmetric. Because the incumbent telephone company was expected to have a higher value than others, the Vickrey auction was a poor choice for that reason as well. Some smaller licenses sold for tenths of 1% of the highest bid.

In a private values setting, the Vickrey auction and the English auction are much easier on bidders than a regular sealed-bid auction because of the dominant strategy. The sealed-bid auction requires bidders to forecast their rivals' likely bids and produces the risks of either bidding more than necessary or losing the bidding. Thus, the regular sealed-bid auction has undesirable properties. Moreover, bidders in the sealed-bid auction have an incentive to bribe the auctioneer to reveal the best bid by rivals because that is useful information in formulating a bid. Such (illegal) bribery occurs from time to time in government contracting.

On the other hand, the regular sealed-bid auction has an advantage over the other two because it makes price fixing more difficult. A bidder can cheat on a conspiracy and not be detected until after the current auction is complete.

Another disadvantage of the sealed-bid auction is that it is easier to make certain kinds of bidding errors. In the U.S. PCS auctions, in which rights to use the radio spectrum for cellular phones was sold for around $20 billion, one bidder, intending to bid $200,000, inadvertently bid $200,000,000. Such an error isn't possible in an English auction because prices rise at a measured pace. And such

an error has little consequence in a Vickrey auction because getting the price wrong by an order of magnitude requires two bidders to make such errors.

Key Takeaways

- There can be more than one way of implementing a given kind of auction.
- The Vickrey auction is a sealed-bid auction where the high bidder wins but pays the second-highest bid. The Vickrey auction is also called a second-price auction: the price is not the highest bid but the second-highest bid.
- The Vickrey auction underlies eBay because when a bidder submits a bid in the eBay auction, the current “going” price is not the highest bid, but the second-highest bid plus a bid increment. Thus, up to the granularity of the bid increment, the basic eBay auction is a Vickrey auction run over time.
- In the private values setting, bidders in a Vickrey auction have a dominant strategy to bid their value. The Vickrey auction is revenue equivalent to the other three auctions.
- Because the Vickrey auction induces bidders to bid their value, it is said to be demand revealing.

This page titled 20.4: Vickrey Auction is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

20.5: The Winner’s Curse and Linkage 20.5: The Winner’s Curse and LinkageEdit Update titlecancel

Learning Objectives

- How do I interpret the bids of others when other bidders may have relevant information about the value of the good?

I paid too much for it, but it's worth it.

-Sam Goldwyn

The analysis so far has been conducted under the restrictive assumption of private values. In most contexts, bidders are not sure of the actual value of the item being sold, and information held by others is relevant to the valuation of the item. If I estimate an antique to be worth $5,000, but no one else is willing to bid more than $1,000, I might revise my estimate of the value down. This revision leads bidders to learn from the auction itself what the item is worth.

The early bidders in the sale of oil lease rights in the Gulf of Mexico (the outer continental shelf) were often observed to pay more than the rights were worth. This phenomenon came to be known as the **winner’s curse**. The winner’s curse is the fact that *the bidder who most overestimates the value of the object wins the bidding.*

Naïve bidders who don’t adjust for the winner’s curse tend to lose money because they win the bidding only when they’ve bid too high.

Figure 20.1 Normally Distributed Estimates

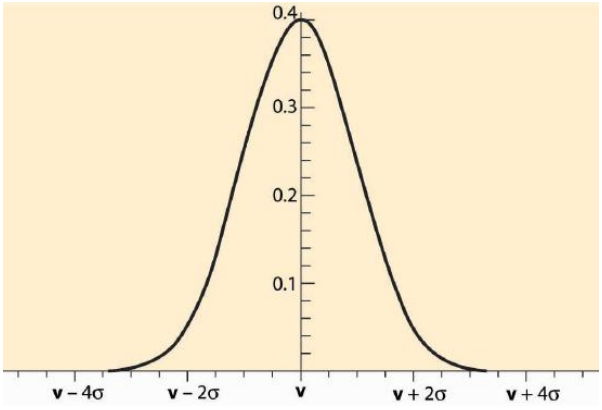


Figure 20.1. The estimates are correct on average, which is represented by the fact that the distribution is centered on the true value *v*. Thus, a randomly chosen bidder will have an estimate that is too high as often as it is too low, and the average estimate of a randomly selected bidder will be correct. However, the winner of an auction will tend to be the bidder with the highest estimate, not a randomly chosen bidder. The highest of five bidders will have an estimate that is too large 97% of the time. The only way the highest estimate is not too large is if all the estimates are below the true value. With 10 bidders, the highest estimate is larger than the true value with probability 99.9% because the odds that all the estimates are less than the true value is $(1/2)^{10} = 0.1\%$. This phenomenon—that auctions tend to select the bidder with the highest estimate, and the highest estimate is larger than the true value most of the time—is characteristic of the winner’s curse.

A savvy bidder corrects for the winner’s curse. Such a correction is actually quite straightforward when a few facts are available, and here a simplified presentation is given. Suppose there are *n* bidders for a common value good, and the bidders receive normally distributed estimates that are correct on average. Let σ be the standard deviation of the estimates.The standard deviation is a measure of the dispersion of the distribution and is the square root of the average of the square of the difference of the random value and its mean. The estimates are also assumed to be independently distributed around the true value. Note that estimating the mean adds an additional layer of complexity. Finally, suppose that no prior information is given about the likely value of the good.

In this case, it is a straightforward matter to compute a correction for the winner’s curse. Because the winning bidder will generally be the bidder with the highest estimate of value, the winner’s curse correction should be the expected amount by which the highest value exceeds the average value. This can be looked up in a table for the normal distribution. The values are given for selected numbers *n* in Table 20.1. This table shows, as a function of the number of bidders, how much each bidder should reduce his estimate of value to correct for the fact that auctions select optimistic bidders. The units are standard deviations.

Table 20.1 Winner’s Curse Correction

n	1	2	3	4	5	10	15
WCC (σ)	0	.56	.85	1.03	1.16	1.54	1.74
n	20	25	50	100	500	1000	10,000
WCC (σ)	1.87	1.97	2.25	2.51	3.04	3.24	3.85

For example, with one bidder, there is no correction because it was supposed that the estimates are right on average. With two bidders, the winner's curse correction is the amount that the higher of two will be above the mean, which turns out to be 0.56σ, a little more than half a standard deviation. This is the amount that should be subtracted from the estimate to ensure that, when the bidder wins, the estimated value is correct, on average. With four bidders, the highest is a bit over a whole standard deviation. As is apparent from the table, the winner's curse correction increases relatively slowly after 10 or 15 bidders. With a million bidders, it is 4.86σ.

The standard deviation σ measures how much randomness or noise there is in the estimates. It is a measure of the average difference between the true value and the estimated value, and thus the average level of error. Oil companies know from their history of estimation how much error arises in the company estimates. Thus, they can correct their estimates to account for the winner's curse using their historical inaccuracies.

Bidders who are imperfectly informed about the value of an item for sale are subject to losses arising from the way auctions select the winning bidder. The winning bidder is usually the bidder with the highest estimate, and that estimate is too high on average. The difference between the highest estimate and the average estimate is known as the winner's curse correction. The size of the winner's curse correction is larger the more bidders there are, but it tends to grow slowly beyond a dozen or so bidders.

If the bidders have the same information on a common value item, they will generally not earn profits on it. Indeed, there is a general principle that it is the privacy of information, rather than the accuracy of information, that leads to profits. Bidders earn profits on the information that they hold that is not available to others. Information held by others will be built into the bid price and therefore not lead to profits.

The U.S. Department of the Interior, when selling offshore oil leases, not only takes an up-front payment (the winning bid) but also takes one-sixth of the oil that is eventually pumped. Such a royalty scheme links the payment made to the outcome and, in a way, shares risk because the payment is higher when there is more oil. Similarly, a book contract provides an author with an upfront payment and a royalty. Many U.S. Department of Defense (DOD) purchases of major weapons systems involve cost-sharing, where the payments made pick up a portion of the cost. Purchases of ships, for example, generally involve 50%–70% cost sharing, which means the DOD pays a portion of cost overruns. The contract for U.S. television broadcast rights for the Summer Olympics in Seoul, South Korea, involved payments that depended on the size of the U.S. audience.

Royalties, cost-sharing, and contingent payments generally link the actual payment to the actual value, which is unknown at the time of the auction. Linkage shares risk, but linkage does something else, too. Linkage reduces the importance of estimates in the auction, replacing the estimates with actual values. That is, the price a bidder pays for an object, when fully linked to the true value, is just the true value. Thus, linkage reduces the importance of estimation in the auction by taking the price out of the bidder's hands, at least partially.

The **linkage principle**The linkage principle, and much of modern auction theory, was developed by Paul Milgrom (1948–). states that in auctions where bidders are buyers, the expected price rises the more the price is linked to the actual value. (In a parallel fashion, the expected price in an auction where bidders are selling falls.) Thus, linking price to value generally improves the performance of auctions. While this is a mathematically deep result, an extreme case is straightforward to understand. Suppose the government is purchasing by auction a contract for delivery of 10,000 gallons of gasoline each week for the next year. Suppliers face risk in the form of gasoline prices; if the government buys at a fixed price, the suppliers' bids will build in a cushion to compensate for the risk and for the winner's curse. In addition, because their estimates of future oil prices will generally vary, they will earn profits based on their private information about the value. In contrast, if the government buys only delivery and then pays for the cost of the gasoline, whatever it might be, any profits that the bidders earned based on their ability to estimate gasoline prices evaporate. The overall profit level of bidders falls, and the overall cost of the gasoline supply can fall. Of course, paying the cost of the gasoline reduces the incentive of the supplier to shop around for the best price, and that agency incentive effect must be balanced against the reduction in bidder profits from the auction to select a supplier.

Key Takeaways

- Auctions, by their nature, select optimistic bidders. This phenomenon—that auctions tend to select the bidder with the highest estimate, and the highest estimate is larger than the true value most of the time—is known as the winner's curse.
- A savvy bidder corrects for the winner's curse.
- The size of the winner's curse correction is larger the more bidders there are, but it tends to grow slowly beyond a dozen or so bidders.
- There is a general principle that it is the privacy of information, rather than the accuracy of information, that leads to profits. Information held by others will be built into the bid price and therefore not lead to profits.
- The linkage principle states that in auctions where bidders are buyers, the expected price rises the more the price is linked to the actual value. Examples of linkage include English and Vickrey auctions, which link the price to the second bidder's information, and the use of royalties or cost shares.

This page titled 20.5: The Winner's Curse and Linkage is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

20.6: Auction Design 20.6: Auction DesignEdit Update titlecancel

Learning Objectives

- What kind of auction should I hold to sell something?
- Should I impose a minimum bid?
- Should I use an open- or sealed-bid auction?

We saw in Section 20.5 that the English auction tends to reduce regret relative to sealed-bid auctions and that the linkage principle suggests tying payments to value where possible. These are examples of auction design, in which auctions are designed to satisfy objectives of the auction designer. Proper auction design should match the rules of the auction to the circumstances of the bidders and the goal of the seller. Some of the principles of auction design include:

- Impose an appropriate reserve price or minimum bid
- Use ascending price (English) auctions rather than sealed-bid auctions
- Reveal information about the value of the item
- Conceal information about the extent of competition
- Handicap bidders with a known advantage

However, many of these principles change if the seller is facing a cartel. For example, it is easier for bidders to collude in a sealed-bid auction than in an English auction, and reserve prices should be made relatively high.

Reserve prices (minimum bids) have several effects. They tend to force marginal bidders to bid a bit higher, which increases the bids of all bidders and thus reduces bidder profits. However, reserve prices also lead to a failure to sell on occasion, and the optimal reserve trades off this failure to sell against the higher prices. In addition, reserve prices may reduce the incentive of bidders to investigate the sale, thus reducing participation, which is an additional negative consideration for a high reserve price.

Ascending price auctions like the English auction have several advantages. Such auctions reduce the complexity of the bidder's problem because bidders can stretch their calculations out over time and because bidders can react to the behavior of others and not plan for every contingency in advance. In addition, because bidders in an English auction can see the behavior of others, there is a linkage created—the price paid by the winning bidder is influenced not just by that bidder's information but also by the information held by others, tending to drive up the price, which is an advantage for the seller.

One caveat to the selection of the English auction is that risk aversion doesn't affect the outcome in the private values case. In contrast, in a sealed-bid auction, risk aversion works to the advantage of the seller because bidders bid a little bit higher than they would have otherwise to reduce the risk of losing. Thus, in the private values case, risk-averse bidders will bid higher in the sealed-bid auction than in the English auction.

When considering the revelation of information, there is always an issue of lying and misleading. In the long run, lying and misleading are found out, and thus the standard approach is to ignore the possibility of lying. Making misleading statements is, in the long run, the same thing as silence because those who repeatedly lie or mislead are eventually discovered and then not believed. Thus, in the long run, a repeat seller has a choice of being truthful or silent. Because of the linkage principle, the policy of revealing truthful information about the value of the good for sale dominates the policy of concealing information because the revelation of information links the payment to the actual outcome.

In contrast, revealing information about the extent of competition may not increase the prices. Consider the case where occasionally there are three bidders, or the case where this is only one. If the extent of competition is concealed, bidders will bid without knowing the extent of competition. If the bidders are risk neutral, it turns out that the revelation doesn't matter and the outcomes are the same on average. If, in contrast, bidders are risk averse, the concealment of information tends to increase the bid prices because the risk created by the uncertainty about the extent of competition works to the advantage of the seller. Of course, it may be difficult to conceal the extent of competition in the English auction, suggesting that a sealed-bid auction should be used instead.

Bidders with a large, known advantage have several deleterious effects. For example, incumbent telephone companies generally are willing to pay more for spectrum in their areas than outsiders are. Bidders bidding at an advantage discourage the participation of others because the others are likely to lose. This can result in a bidder with an advantage facing no competition and picking up the good cheaply. Second, rivals don't present much competition to the advantaged bidder, even if the rivals do participate.

Consequently, when a bidder has a large advantage over rivals, it is advantageous to handicap the advantaged bidder, thus favoring the rivals. This handicapping encourages participation and levels the playing field, forcing the advantaged bidder to bid more competitively to win.

A common means of favoring disadvantaged bidders is by the use of **bidder credits**. For example, with a 20% bidder credit for disadvantaged bidders, a disadvantaged bidder has to pay only 80% of the face amount of the bid. This lets such a bidder bid 25% more (because a $100 payment corresponds to a $125 bid) than she would have otherwise, which makes the bidder a more formidable competitor. Generally, the ideal bidder credit is less than the actual advantage of the advantaged bidder.

Auction design is an exciting development in applied industrial organization, in which economic theory and experience is used to improve the performance of markets. The U.S. Federal Communications auctions of spectrum were the first major instance of auction design in an important practical setting, and the auction design was credited with increasing substantially the revenue raised by the government.

Key Takeaways

- Some of the principles of auction design include:
  - Impose an appropriate reserve price or minimum bid
  - Use ascending price (English) auctions rather than sealed-bid auctions
  - Reveal information about the value of the item
  - Conceal information about the extent of competition
  - Handicap bidders with a known advantage
- The optimal reserve trades off this failure to sell against the higher prices when sales arise.
- Ascending price auctions create linkage and reduce the complexity of the bidder's problem.
- Consistent revelation of accurate information about the value of a good increases average prices through linkage, relative to the policy of concealing information.
- Revealing information about the extent of competition may not increase the prices.
- When a bidder has a large advantage over rivals, it is advantageous to handicap the advantaged bidder, favoring the rivals. This handicapping encourages participation and levels the playing field, forcing the advantaged bidder to bid more competitively to win.
- A common means of favoring disadvantaged bidders is by the use of bidder credits.
- Auction design is used to improve the performance of markets and is becoming a field in its own right.

This page titled 20.6: Auction Design is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

21: Antitrust

21.1: Sherman Act
21.2: Clayton Act
21.3: Price Fixing
21.4: Mergers

This page titled 21: Antitrust is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

21.1: Sherman Act 21.1: Sherman ActEdit Update titlecancel

Learning Objectives

- What is the first U.S. antitrust law?
- What is antitrust anyway?

In archaic language, a **trust** (which is now known as a cartel) was a group of firms acting in concert. The antitrust laws that made such trusts illegal were intended to protect competition. In the United States, these laws are enforced by the **U.S. Department of Justice**'s (DOJ) Antitrust Division and by the Federal Trade Commission (FTC). The United States began passing laws during a time when some European nations were actually passing laws forcing firms to join industry cartels. By and large, however, the rest of the world has since copied the U.S. antitrust laws in one form or another.

The **Sherman Act**, passed in 1890, was the first significant piece of antitrust legislation. It has two main requirements.

Section 1. Trusts, etc., in restraint of trade illegal; penalty

Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is declared to be illegal. Every person who shall make any contract or engage in any combination or conspiracy hereby declared to be illegal shall be deemed guilty of a felony, and, on conviction thereof, shall be punished by fine not exceeding $10,000,000 if a corporation, or, if any other person, $350,000, or by imprisonment not exceeding 3 years, or by both said punishments, in the discretion of the court.

Section 2. Monopolizing trade a felony; penalty

Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a felony, and, on conviction thereof, shall be punished by fine not exceeding $10,000,000 if a corporation, or, if any other person, $350,000, or by imprisonment not exceeding 3 years, or by both said punishments, in the discretion of the court.The current fines were instituted in 1974; the original fines were $5,000, with a maximum imprisonment of one year. The Sherman Act is 15 U.S.C. § 1.

The phrase *in restraint of trade* is challenging to interpret. Early enforcement of the Sherman Act followed the Peckham Rule, named for noted Justice Rufus Peckham, which interpreted the Sherman Act to prohibit contracts that reduced output or raised prices while permitting contracts that would increase output or lower prices. In one of the most famous antitrust cases ever, the United States sued Standard Oil, which had monopolized the transportation of oil from Pennsylvania to the East Coast cities of the United States in 1911.

The exact meaning of the Sherman Act had not been settled at the time of the Standard Oil case. Indeed, Supreme Court Justice Edward White suggested that, because contracts by their nature set the terms of trade and thus restrain trade to those terms, and Section 1 makes contracts restraining trade illegal, one could read the Sherman Act to imply that *all* contracts were illegal. Chief Justice White concluded that, because Congress couldn't have intended to make all contracts illegal, the intent must have been to make unreasonable contracts illegal, and he therefore concluded that judicial discretion is necessary in applying the antitrust laws. In addition, Chief Justice White noted that the act makes monopolizing illegal, but doesn't make having a monopoly illegal. Thus, Chief Justice White interpreted the act to prohibit certain acts leading to monopoly, but not monopoly itself.

The legality of monopoly was further clarified through a series of cases, starting with the 1945 Alcoa case, in which the United States sued to break up the aluminum monopoly Alcoa. The modern approach involves a **two-part test**. First, does the firm have monopoly power in a market? If it does not, no monopolization has occurred and there is no issue for the court. Second, if it does, did the firm use illegal tactics to extend or maintain that monopoly power? In the language of a later decision, did the firm engage in "the willful acquisition or maintenance of that power as distinguished from growth or development as a consequence of superior product, business acumen or historic accident" (*U.S. v. Grinnell*, 1966)?

There are several important points that are widely misunderstood and even misreported in the press. First, the Sherman Act does not make having a monopoly illegal. Indeed, three legal ways of obtaining a monopoly—a better product, running a better business, or luck—are mentioned in one decision. It is illegal to leverage an existing monopoly into new products or services, or to engage in anticompetitive tactics to maintain the monopoly. Moreover, you must have monopoly power currently to be found guilty of illegal tactics.

When the DOJ sued Microsoft over the incorporation of the browser into the operating system and other acts (including contracts with manufacturers prohibiting the installation of Netscape), the allegation was not that Windows was an illegal monopoly. The DOJ alleged Microsoft was trying to use its Windows monopoly to monopolize another market, the Internet browser market. Microsoft's defense was twofold. First, it claimed not to be a monopoly, citing the 5% share of Apple. (Linux had a negligible share at the time.) Second, it alleged that a browser was not a separate market but an integrated product necessary for the functioning of the operating system. This defense follows the standard two-part test.

Microsoft's defense brings up the question, What is a monopoly? The simple answer to this question depends on whether there are good substitutes in the minds of consumers, so that they may substitute an alternate product in the event of bad behavior by the seller. By this test, Microsoft had an operating system monopoly; in spite of the fact that there was a rival product, Microsoft could increase the price, tie the browser and MP3 player to the operating system, or even disable Word Perfect, and most consumers would not switch to the competing operating system. However, Microsoft's second defense, that the browser wasn't a separate market, was a much more challenging defense to assess.

The Sherman Act provides criminal penalties, which are commonly applied in price-fixing cases—that is, when groups of firms join together and collude to raise prices. Seven executives of General Electric (GE) and Westinghouse, who colluded in the late 1950s to set the prices of electrical turbines, each spent several years in jail, and incurred over $100 million in fines. In addition, Archer Daniels Midland executives went to jail after their 1996 conviction for fixing the price of lysine, which approximately doubled the price of this common additive to animal feed. When highway contractors are convicted of bid-rigging, the conviction is typically under the Sherman Act for monopolizing their market.

KEY TAKEAWAYS

- A trust (now known as a cartel) is a group of firms acting in concert. The antitrust laws made such trusts illegal and were intended to protect competition. In the United States, these laws are enforced by the Department of Justice's (DOJ) Antitrust Division and by the Federal Trade Commission (FTC).
- The Sherman Act, passed in 1890, is the first significant piece of antitrust legislation. It prevents mergers and cartels that would increase prices.
- Having a monopoly is legal, provided it is obtained through legal means. Legal means include "superior product, business acumen or historic accident."
- Modern antitrust investigations involve a two-part test. First, does the firm have monopoly power in a market? If it does not, no monopolization has occurred and there is no issue for the court. If it does, did the firm use illegal tactics to extend or maintain that monopoly power?
- The Sherman Act provides criminal penalties, which are commonly applied in price-fixing cases.

This page titled 21.1: Sherman Act is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

21.2: Clayton Act 21.2: Clayton ActEdit Update titlecancel

Learning Objectives

- What other major antitrust legislation exists in the United States?
- What is predatory pricing and why is it illegal?
- Is price discrimination illegal?

Critics of the Sherman Act, including famous trust-buster President Teddy Roosevelt, felt the ambiguity of the Sherman Act was an impediment to its use and that the United States needed a more detailed law setting out a list of illegal activities. The **Clayton Act**, 15 U.S.C. §§ 12–27, was passed in 1914 and it adds detail to the Sherman Act. The same year, the FTC Act was passed, creating the **Federal Trade Commission (FTC)**, which has authority to enforce the Clayton Act as well as to engage in other consumer protection activities.

The Clayton Act does not have criminal penalties, but it does allow for monetary penalties that are three times as large as the damage created by the illegal behavior. Consequently, a firm, motivated by the possibility of obtaining a large damage award, may sue another firm for infringement of the Clayton Act. A plaintiff must be directly harmed to bring such a suit. Thus, customers who paid higher prices or firms that were driven out of business by exclusionary practices are permitted to sue under the Clayton Act. When Archer Daniels Midland raised the price of lysine, pork producers who bought lysine would have standing to sue, but final pork consumers who paid higher prices for pork, but who didn't directly buy lysine, would not.

Highlights of the Clayton Act include:

- Section 2, which prohibits price discrimination that would lessen competition
- Section 3, which prohibits exclusionary practices, such as tying, exclusive dealing, and predatory pricing, that lessen competition
- Section 7, which prohibits share acquisition or merger that would lessen competition or create a monopoly

The language *lessen competition* is generally understood to mean that a significant price increase becomes possible; that is, competition has been harmed if the firms in the industry can successfully increase prices.

Section 2 is also known as **Robinson-Patman** because of a 1936 amendment by that name. It prohibits price discrimination that lessens competition. Thus, price discrimination to final consumers is legal under the Clayton Act; the only way price discrimination can lessen competition is if one charges different prices to different businesses. The logic of the law was articulated in the 1948 Morton Salt decision, which concluded that lower prices to large chain stores gave an advantage to those stores, thus injuring competition in the grocery store market. The discounts in that case were not cost-based, and it is permissible to charge different prices based on costs.

Section 3 rules out practices that lessen competition. A manufacturer who also offers service for the goods it sells may be prohibited from favoring its own service organization. Generally manufacturers may not require the use of the manufacturer's own service. For example, an automobile manufacturer can't require the use of replacement parts made by the manufacturer, and many car manufacturers have lost lawsuits on this basis. In an entertaining example, Mercedes prohibited Mercedes dealers from buying Bosch parts directly from Bosch, even though Mercedes itself was selling Bosch parts to the dealers. This practice was ruled illegal because the quality of the parts was the same as Mercedes's (indeed, identical), so Mercedes's action lessened competition.

Predatory pricing involves pricing below cost in order to drive a rival out of business. It is relatively difficult for a firm to engage in predation simply because it only makes sense if, once the rival is eliminated, the predatory firm can then increase its prices and recoup the losses incurred. The problem is that once the prices go up, entry becomes attractive; so what keeps other potential entrants away? One answer is reputation: a reputation for a willingness to lose money in order to dominate the market could deter potential entrants. Like various rare diseases that happen more often on television shows than in the real world (e.g., Tourette's syndrome), predatory pricing probably happens more often in textbooks than in the real world.Economists have argued that American Tobacco, Standard Oil, and AT&T each engaged in predation in their respective industries.

The FTC also has authority to regulate mergers that would lessen competition. As a practical matter, the DOJ and the FTC divide responsibility for evaluating mergers. In addition, other agencies may also have jurisdiction over mergers and business tactics. The Department of Defense has oversight of defense contractors, using a threat of "we're your only customer." The Federal

Communications Commission has statutory authority over telephone and television companies. The Federal Reserve Bank has authority over national and most other banks.

Most states have antitrust laws as well, and they can challenge mergers that would affect commerce in the respective state. In addition, attorneys general of many states may join the DOJ or the FTC in suing to block a merger or in other antitrust actions, or they can sue independently. For example, many states joined the Department of Justice in its lawsuit against Microsoft. Forty-two states jointly sued the major record companies over their "minimum advertised prices (MAP)" policies, which the states argued resulted in higher compact disc prices. The MAP case settlement resulted in a modest payment to compact disc purchasers. The FTC had earlier extracted an agreement to stop the practice.

Key Takeaways

- The Clayton Act was passed in 1914 and adds detail to the Sherman Act. The FTC, which has authority to enforce the Clayton Act, as well as engage in other consumer protection activities, was created the same year.
- The Clayton Act does not have criminal penalties, but it does allow for monetary penalties that are three times as large as the damage created by the illegal behavior.
- Highlights of the Clayton Act include:
  - Section 2, which prohibits price discrimination that would lessen competition
  - Section 3, which prohibits exclusionary practices, such as tying, exclusive dealing, and predatory pricing, that lessen competition
  - Section 7, which prohibits share acquisition or merger that would lessen competition or create a monopoly
- The language *lessen competition* is generally understood to mean that a significant price increase becomes possible; that is, competition has been harmed if the firms in the industry can successfully increase prices.
- Predatory pricing involves pricing below cost in order to drive a rival out of business.
- The DOJ and the FTC divide responsibility for evaluating mergers.
- Most states have antitrust laws as well, and they can challenge mergers that would affect commerce in the respective state.

This page titled 21.2: Clayton Act is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

21.3: Price Fixing 21.3: Price FixingEdit Update titlecancel

Learning Objectives

- What is price fixing and how does it work?

Price fixing, which is called **bid-rigging** in a bidding context, involves a group of firms agreeing to increase the prices they charge and restrict competition against each other. The most famous example of price fixing is probably the Great Electrical Conspiracy in which GE and Westinghouse (and a smaller firm, Allis-Chalmers) fixed the prices of turbines used for electricity generation. Generally these turbines were the subject of competitive (or, in this case, not-so-competitive) bidding, and the companies set the prices by designating a winner for each bidding situation and using a price book to provide identical bids by all companies. An amusing element of the price-fixing scheme was the means by which the companies identified the winner in any given competition: they used the phase of the moon. The phase of the moon determined the winner, and each company knew what to bid based on the phase of the moon. Executives from the companies met often to discuss the terms of the price-fixing arrangement, and the Department of Justice (DOJ) acquired a great deal of physical evidence in the process of preparing its 1960 case. Seven executives went to jail and hundreds of millions of dollars in fines were paid.

Most convicted price-fixers are from small firms. The turbine conspiracy and the Archer Daniels Midland lysine conspiracy are unusual. (There is evidence that large vitamin manufacturers conspired in fixing the price of vitamins in many nations of the world.) Far more common conspiracies involve highway and street construction firms, electricians, water and sewer construction companies, or other owner-operated businesses. Price fixing seems most common when owners are also managers and there are a small number of competitors in a given region.

As a theoretical matter, it should be difficult for a large firm to motivate a manager to engage in price fixing. The problem is that the firm can't write a contract promising the manager extraordinary returns for successfully fixing prices because such a contract itself would be evidence and moreover implicate higher management. Indeed, Archer Daniels Midland executives paid personal fines of $350,000, and each served 2 years in jail. Thus, it is difficult to offer a substantial portion of the rewards of price fixing to managers in exchange for the personal risks the managers would face from engaging in price fixing. Most of the gains of price fixing accrue to shareholders of large companies, while large risks and costs fall on executives. In contrast, for smaller businesses in which the owner is the manager, the risks and rewards are borne by the same person, and thus the personal risk is more likely to be justified by the personal return.

We developed earlier a simple theory of cooperation, in which the grim trigger strategy was used to induce cooperation. Let us apply that theory to price fixing. Suppose that there are *n* firms and that they share the monopoly profits π_m equally if they collude. If one firm cheats, that firm can obtain the entire monopoly profits until the others react. This is clearly the most the firm could get from cheating. Once the others react, the collusion breaks down and the firms earn zero profits (the competitive level) from then on. The cartel is feasible if 1/*n* of the monopoly profits forever is better than the whole monopoly profits for a short period of time. Thus, cooperation is sustainable if $\Pi mn(1-\delta) = \Pi mn(1+\delta+\delta 2+...) \geq nm$.

The left-hand side of the equation gives the profits from cooperating—the present value of the 1/*n* share of the monopoly profits. In contrast, if a firm chooses to cheat, it can take at most the monopoly profits, but only temporarily. How many firms will this sustain? The inequality simplifies to $n \leq 11-\delta$. Suppose the annual interest rate is 5% and the reaction time is 1 week—that is, a firm that cheats on the cooperative agreement sustains profits for a week, after which time prices fall to the competitive level. In this case, 1 − δ is a week's worth of interest (δ is the value of money received in a week), and therefore $\delta = 0.95152 = 0.999014$. According to standard theory, the industry with a weeklong reaction time should be able to support cooperation with up to a thousand firms.

There are numerous and varied reasons why this theory fails to work very well empirically, including that some people are actually honest and do not break the law, but we will focus on one game-theoretic reason here. The cooperative equilibrium is not the only equilibrium, and there are good reasons to think that full cooperation is unlikely to persist. The problem is the prisoner's dilemma itself: generally the first participant to turn in the conspiracy can avoid jail. Thus, if one member of a cartel is uncertain whether the other members of a price-fixing conspiracy are contacting the DOJ, that member may race to the DOJ—the threat of one confession may cause them all to confess in a hurry. A majority of the conspiracies that are prosecuted arise because someone—a member who feels guilty, a disgruntled ex-spouse of a member, or perhaps a member who thinks another member is suffering

pangs of conscience—turns them in. Lack of confidence in the other members creates a self-fulfilling prophecy. Moreover, cartel members should lack confidence in the other cartel members who are, after all, criminals.

On average, prosecuted conspiracies were about 7 years old when they were caught. Thus, there is about a 15% chance annually of a breakdown of a conspiracy, at least among those that are eventually caught.

Key Takeaways

- Price fixing, which is called bid rigging in a bidding context, involves a group of firms agreeing to increase the prices they charge and restrict competition against each other.
- The most famous example of price fixing is probably the Great Electrical Conspiracy in which GE and Westinghouse fixed the prices of turbines. The companies used the phase of the moon to determine the winner of government procurement auctions.
- Theoretically, collusions should be easy to sustain; in practice, it does not seem to be.

This page titled 21.3: Price Fixing is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

21.4: Mergers 21.4: MergersEdit Update titlecancel

Learning Objectives

- How does the government decide which mergers to block and which to permit?

The U.S. Department of Justice (DOJ) and the Federal Trade Commission (FTC) share responsibility for evaluating mergers. Firms with more than $50 million in assets are required under the Hart-Scott-Rodino Act to file with the government an intention to merge with another firm. The government then has a limited amount of time to either approve the merger or request more information (called a second request). Once the firms have complied with the second request, the government again has a limited amount of time before it either approves the merger or sues to block it. The government agencies themselves don't stop the merger, but instead they sue to block the merger, asking a federal judge to prevent the merger as a violation of one of the antitrust laws. Mergers are distinct from other violations because they have not yet occurred at the time the lawsuit is brought, so there is no threat of damages or criminal penalties; the only potential penalty imposed on the merging parties is that the proposed merger may be blocked.

Many proposed mergers result in settlements. As part of the settlement associated with GE's purchase of Radio Corporation of America (RCA) in 1986, a small appliance division of GE's was sold to Black & Decker, thereby maintaining competition in the small kitchen appliance market. In the 1999 merger of oil companies Exxon and Mobil, a California refinery, shares in oil pipelines connecting the Gulf with the Northeast, and thousands of gas stations were sold to other companies. The 1996 merger of Kimberley-Clark and Scott Paper would have resulted in a single company with over 50% of the facial tissue and baby wipes markets, and in both cases divestitures of production capacity and the Scotties brand name preserved competition in the markets. Large bank mergers, oil company mergers, and other large companies usually present some competitive concerns, and the majority of these cases are solved by divestiture of business units to preserve competition.

A **horizontal merger** is a merger of competitors, such as Exxon and Mobil or two banks located in the same city. In contrast, a **vertical merger** is a merger between an input supplier and input buyer. The attempt by book retailer Barnes and Noble to purchase the intermediary Ingram, a company that buys books from publishers and sells to retailers but doesn't directly sell to the public, would have resulted in a vertical merger. Similarly, Disney is a company that sells programs to television stations (among other activities), so its purchase of TV network ABC was a vertical merger. The AOL-Time Warner merger involved several vertical relationships. For example, Time Warner is a large cable company, and cable represents a way for AOL to offer broadband services. In addition, Time Warner is a content provider, and AOL delivers content to Internet subscribers.

Vertical mergers raise two related problems: foreclosure and raising rivals' costs. **Foreclosure** refers to denying access to necessary inputs. Thus, the AOL-Time Warner merger threatened rivals to AOL Internet service (like EarthLink) with an inability to offer broadband services to consumers with Time Warner cable. This potentially injures competition in the Internet service market, forcing Time Warner customers to use AOL. In addition, by bundling Time Warner content and AOL Internet service, users could be forced to purchase AOL Internet service in order to have access to Time Warner content. Both of these threaten foreclosure of rivals, and both were resolved to the government's satisfaction by promises that the merged firm would offer equal access to rivals.

Raising rivals' costs is a softer version of foreclosure. Rather than deny access to content, AOL Time Warner could instead make the content available under disadvantageous terms. For example, American Airlines developed the Sabre computerized reservation system, which was used by about 40% of travel agents. This system charged airlines, rather than travel agents, for bookings. Consequently, American Airlines had a mechanism for increasing the costs of its rivals: by increasing the price of bookings on the Sabre system. The advantage to American Airlines was not just increased revenue of the Sabre system but also the hobbling of airline rivals. Similarly, banks offer free use of their own automated teller machines (ATMs), but they charge the customers of other banks. Such charges raise the costs of customers of other banks, thus making other banks less attractive and providing an advantage in the competition for bank customers.

The DOJ and the FTC periodically issue horizontal merger guidelines, which set out how mergers will be evaluated. This is a three-step procedure for each product that the merging companies have in common.

The procedure starts by identifying product markets. To identify a product market, start with a product or products produced by both companies. Then ask if the merged parties can profitably raise price by a small but significant and nontransitory increase in price, also known as a **SSNIP** (pronounced "snip"). A SSNIP is often taken to be a 5% price increase, which must prevail for several years. If the companies can profitably increase price by a SSNIP, then they are judged to have monopoly power and

consumers will be directly harmed by the merger. (This is known as a **unilateral effect** because the merging parties will increase price unilaterally after the merger is consummated.) If they can't increase prices, then an additional product has to be added to the group; generally the best substitute is added. Ask whether a hypothetical monopoly seller of these three products can profitably raise price. If it can, an antitrust market has been identified; if it cannot, yet another substitute product must be added. The process stops adding products when enough substitutes have been identified that, if controlled by a hypothetical monopoly, would have their prices significantly increased.

The logic of product market definition is that, if a monopoly wouldn't increase price in a meaningful way, then there is no threat to consumers—any price increase won't be large or won't last. The market is defined by the smallest set of products for which consumers can be harmed. The test is also known as the hypothetical monopoly test.

The second step is to identify a geographic market. The process starts with an area in which both companies sell and asks if the merged company has an incentive to increase price by a SSNIP. If it does, that geographic area is a geographic market. If it does not, it is because buyers are substituting outside the area to buy cheaply, and the area must be expanded. For example, owning all the gas stations on a corner doesn't let one increase price profitably because an increase in price leads to substitution to gas stations a few blocks away. If one company owned all the stations in a half-mile radius, would it be profitable to increase price? Probably not because there would still be significant substitution to more distant stations. Suppose, instead, that one owned all the stations for a 15-mile radius. Then an increase in price in the center of the area is not going to be thwarted by too much substitution outside the area, and the likely outcome is that prices would be increased by such a hypothetical monopoly. In this case, a geographic market has been identified. Again, parallel to the product market definition, a geographic market is the smallest area in which competitive concerns would be raised by a hypothetical monopoly. In any smaller area, attempts to increase price are defeated by substitution to sellers outside the area.

The product and geographic markets together are known as a **relevant antitrust market** (relevant for the purposes of analyzing the merger).

The third and last step of the procedure is to identify the level of concentration in each relevant antitrust market. The Hirschman-Herfindahl Index (HHI) is used for this purpose. The HHI is the sum of the squared market shares of the firms in the relevant antitrust market, and it is justified because it measures the price-cost margin in the Cournot model. Generally, in practice, the shares in percentage are used, which makes the scale range from 0 to 10,000. For example, if one firm has 40%, one has 30%, one has 20%, and the remaining firm has 10%, the HHI is $40^2 + 30^2 + 20^2 + 10^2 = 3,000$

Usually, anything over 1,800 is considered very concentrated, and anything over 1,000 is concentrated.

Suppose firms with shares x and y merge, and nothing in the industry changes besides the combining of those shares. Then the HHI goes up by $(x + y)^2 - x^2 - y^2 = 2xy$. This is referred to as the change in the HHI. The merger guidelines suggest that the government will likely challenge mergers with (a) a change of 100 and a concentrated post-merger HHI, or (b) a change of 50 and a very concentrated post-merger HHI. It is more accurate in understanding the merger guidelines to say that the government likely won't challenge unless either (a) or (b) is met. Even if the post-merger HHI suggests a very concentrated industry, the government is unlikely to challenge if the change in the HHI is less than 50.

Several additional factors affect the government's decision. First, if the firms are already engaging in price discrimination, the government may define quite small geographic markets, possibly as small as a single customer. Second, if one firm is very small (less than 1%) and the other not too large (less than 35%), the merger may escape scrutiny because the effect on competition is likely small. Third, if one firm is going out of business, the merger may be allowed as a means of keeping the assets in the industry. Such was the case with Greyhound's takeover of Trailways, a merger that produced a monopoly of the only intercity bus companies in the United States.

Antitrust originated in the United States, and the United States remains the most vigorous enforcer of antitrust laws. However, the European Union has recently taken a more aggressive antitrust stance, and in fact it has blocked mergers that obtained tentative U.S. approval, such as GE and Honeywell.

Antitrust is, in some sense, the applied arm of oligopoly theory. Because real situations are so complex, the application of oligopoly theory to antitrust analysis is often challenging, and we have only scratched the surface of many of the more subtle issues of law and economics in this text. For example, intellectual property, patents, and standards all have their own distinct antitrust issues.

Key Takeaways

- Firms with large assets are required to notify the government prior to merging.
- Many proposed mergers result in settlements.
- A horizontal merger is a merger of competitors. In contrast, a vertical merger is a merger between an input supplier and input buyer.
- Vertical mergers raise two problems: foreclosure and raising rivals' costs. Foreclosure refers to denying access to necessary inputs. Raising rivals' costs is a softer version of foreclosure because it charges more for inputs.
- Mergers are evaluated by a three-step procedure that involves looking at product market, geographic market, and effects.
- A product market is a set of products sufficiently extensive that a monopolist can profitably raise price by a small but significant and nontransitory increase in price, also known as a SSNIP (pronounced "snip").
- The logic of product market definition is that, if a monopoly wouldn't increase price in a meaningful way and there is no threat to consumers, any price increase won't be large or won't last. The market is defined by the smallest set of products for which consumers can be harmed. The test is also known as the hypothetical monopoly test.
- The second step is to identify a geographic market, which exactly parallels the product market, looking for an area large enough that a hypothetical monopolist over the product market in that geographic market would profitably raise price by a SSNIP.
- The product and geographic markets together are known as a relevant antitrust market (relevant for the purposes of analyzing the merger).
- The third and last step of the procedure is to identify the level of concentration in each relevant antitrust market. The Hirschman-Herfindahl Index (HHI) is used for this purpose.
- Several additional factors, including price discrimination and failing firms, affect the government's decision to sue and thus block mergers.
- Antitrust is, in some sense, the applied arm of oligopoly theory.

This page titled 21.4: Mergers is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

Index IndexEdit Update titlecancel

B

bidder credits
20.6: Auction Design

C

Clayton Act
21.2: Clayton Act

corporate finance
11.3: Investment Under Uncertainty

Cournot model
17.1: Cournot Oligopoly

Cournot Oligopoly
17.1: Cournot Oligopoly

D

dominant strategy
16.1: Matrix Games

E

elasticity of demand
3.1: Elasticity

F

folk theorem
16.6: Supergames

G

grim trigger strategy
16.6: Supergames

H

hedonic pricing
1.2: Opportunity Cost

Hotelling model
17.3: Hotelling Differentiation

I

internal rate of return
11.2: Investment

IRR
11.2: Investment

M

matrix payoff game
16.1: Matrix Games

N

Nash equilibrium
16.2: Nash Equilibrium

O

option
11.3: Investment Under Uncertainty

R

reserve prices
20.6: Auction Design

revenue equivalence
20.2: Sealed-bid Auction

S

stage game
16.6: Supergames

supergame
16.6: Supergames

T

total expenditure
3.1: Elasticity

V

Vickrey Auction
20.4: Vickrey Auction

W

winner’s curse
20.5: The Winner’s Curse and Linkage

Index IndexEdit Update titlecancel

> This page is a draft and is under active development.

B

bidder credits
- 20.6: Auction Design

C

Clayton Act
- 21.2: Clayton Act

corporate finance
- 11.3: Investment Under Uncertainty

Cournot model
- 17.1: Cournot Oligopoly

Cournot Oligopoly
- 17.1: Cournot Oligopoly

D

dominant strategy
- 16.1: Matrix Games

E

elasticity of demand
- 3.1: Elasticity

F

folk theorem
- 16.6: Supergames

G

grim trigger strategy
- 16.6: Supergames

H

hedonic pricing
- 1.2: Opportunity Cost

Hotelling model
- 17.3: Hotelling Differentiation

I

internal rate of return
- 11.2: Investment

IRR
- 11.2: Investment

M

matrix payoff game
- 16.1: Matrix Games

N

Nash equilibrium
- 16.2: Nash Equilibrium

O

option
- 11.3: Investment Under Uncertainty

R

reserve prices
- 20.6: Auction Design

revenue equivalence
- 20.2: Sealed-bid Auction

S

stage game
- 16.6: Supergames

supergame
- 16.6: Supergames

T

total expenditure
- 3.1: Elasticity

V

Vickrey Auction
- 20.4: Vickrey Auction

W

winner’s curse
- 20.5: The Winner’s Curse and Linkage

Glossary GlossaryEdit Update titlecancel

absolute advantage | Condition that exists when one production possibilities frontier can produce more of all goods than another.

absolute advantage | Condition that exists when one production possibilities frontier can produce more of all goods than another.

ad valorem | At value.

ad valorem | At value.

arbitrage | Buying and reselling to make a profit.

arbitrage | Buying and reselling to make a profit.

Arrow-Pratt measure of risk aversion | A measure of risk aversion computed as the negative of the ratio of the second derivative of utility divided by the first derivative of utility.

Arrow-Pratt measure of risk aversion | A measure of risk aversion computed as the negative of the ratio of the second derivative of utility divided by the first derivative of utility.

autarky | State of no exchange.

autarky | State of no exchange.

baby boom | A dramatic increase in births for the years 1946 to 1964.

baby boom | A dramatic increase in births for the years 1946 to 1964.

bid increments | The difference between successive price requests.

bid increments | The difference between successive price requests.

bid-rigging | Price fixing in an auction context.

bid-rigging | Price fixing in an auction context.

bidder credits | Favoritism in an auction for certain bidders in the form of discounts on payment.

bidder credits | Favoritism in an auction for certain bidders in the form of discounts on payment.

bliss point | A point that maximizes utility.

bliss point | A point that maximizes utility.

budget line | Set of goods that just exhaust the consumer's budget.

budget line | Set of goods that just exhaust the consumer's budget.

budget set | Set of goods a consumer can afford.

budget set | Set of goods a consumer can afford.

business cycle | Random fluctuations in output.

business cycle | Random fluctuations in output.

capital accounts | Earnings on foreign assets, and the payments from U.S.-based assets owned by foreigners.

capital accounts | Earnings on foreign assets, and the payments from U.S.-based assets owned by foreigners.

certainty equivalent | The amount of money that provides equal utility to the random payoff of the gamble.

certainty equivalent | The amount of money that provides equal utility to the random payoff of the gamble.

ceteris paribus | Latin phrase meaning “other things equal.”

ceteris paribus | Latin phrase meaning “other things equal.”

circle model | A Hotelling model set on a circle.

circle model | A Hotelling model set on a circle.

Civil Rights Act | Legislation that prohibited segregation based on race in schools, public places, and employment.

Civil Rights Act | Legislation that prohibited segregation based on race in schools, public places, and employment.

Coasian bargaining | Solving the problem of externalities through negotiation.

Coasian bargaining | Solving the problem of externalities through negotiation.

Cobb-Douglas production function | A production function that is the product of each input, x, raised to a given power.

Cobb-Douglas production function | A production function that is the product of each input, x, raised to a given power.

common resource | A resource shared by many people.

common resource | A resource shared by many people.

common values | Situation in which bidders in an auction have the same value, which is generally not known with certainty.

common values | Situation in which bidders in an auction have the same value, which is generally not known with certainty.

comparative disadvantage | Condition that exists when one good has a higher opportunity cost over another.

comparative disadvantage | Condition that exists when one good has a higher opportunity cost over another.

comparative static | A prediction that allows one to determine how one variable affects another, at least in the setting described by the model.

comparative static | A prediction that allows one to determine how one variable affects another, at least in the setting described by the model.

compensated demand | Demand that exists when a change in price is accompanied by just enough additional income to keep utility the same.

compensated demand | Demand that exists when a change in price is accompanied by just enough additional income to keep utility the same.

compensating differential | Income or costs that equalize different choices.

compensating differential | Income or costs that equalize different choices.

complement in supply | A good whose cost falls as the amount produced of another good rises.

complement in supply | A good whose cost falls as the amount produced of another good rises.

complements | For a given good x, a good whose consumption increases the value of x.

complements | For a given good x, a good whose consumption increases the value of x.

compound interest | The correct rate of interest that is calculated by adding together the cumulative interest payments earned during one year.

compound interest | The correct rate of interest that is calculated by adding together the cumulative interest payments earned during one year.

concentrated benefits and diffused costs | Situation where a small number of people with strong incentives are able to expropriate a small amount per person from a large number of people.

concentrated benefits and diffused costs | Situation where a small number of people with strong incentives are able to expropriate a small amount per person from a large number of people.

Consolidated annuities or consols | Bonds that pay a fixed amount in perpetuity (forever).

Consolidated annuities or consols | Bonds that pay a fixed amount in perpetuity (forever).

constant elasticity | Condition in which the elasticity remains at the same level while the underlying variables change.

constant elasticity | Condition in which the elasticity remains at the same level while the underlying variables change.

constant returns to scale | Situation that exists when increasing all inputs by the same scalar factor increases output by that scalar factor.

constant returns to scale | Situation that exists when increasing all inputs by the same scalar factor increases output by that scalar factor.

consumer price index (CPI) | A price deflator that adjusts for what it costs to buy a “standard” bundle of food, clothing, housing, electricity, and other items.

consumer price index (CPI) | A price deflator that adjusts for what it costs to buy a “standard” bundle of food, clothing, housing, electricity, and other items.

consumer surplus | The value of consuming a good, minus the price paid.

consumer surplus | The value of consuming a good, minus the price paid.

contract curve | Curve in which every point maximizes one person’s utility given another’s utility.

contract curve | Curve in which every point maximizes one person’s utility given another’s utility.

convex function | The function that lies below the straight line segment connecting two points, for any two points in the interval.

convex function | The function that lies below the straight line segment connecting two points, for any two points in the interval.

corporate finance | The study of funding of operations of companies.

corporate finance | The study of funding of operations of companies.

corporation | A single entity owned by shareholders.

corporation | A single entity owned by shareholders.

Cournot model | A model of imperfect competition where firms simultaneously set quantities.

Cournot model | A model of imperfect competition where firms simultaneously set quantities.

covary | To move together.

covary | To move together.

deadweight loss | The buyer’s values minus the seller’s costs of units that are not economic to trade because of a tax or other interference in the market.

deadweight loss | The buyer's values minus the seller's costs of units that are not economic to trade because of a tax or other interference in the market.

Diminishing marginal returns | The principle that spending more on something reduces the per-unit value produced.

Diminishing marginal returns | The principle that spending more on something reduces the per-unit value produced.

diminishing marginal value | Condition in which the value of the last unit declines as the number consumed rises.

diminishing marginal value | Condition in which the value of the last unit declines as the number consumed rises.

discounted | The practice of adjusting a sequence of payments to account for the interest these payments would earn in financial markets.

discounted | The practice of adjusting a sequence of payments to account for the interest these payments would earn in financial markets.

discretionary expenditures | Government expenditures that aren't transfers.

discretionary expenditures | Government expenditures that aren't transfers.

dominant strategy | A strategy that is optimal no matter what the other players do.

dominant strategy | A strategy that is optimal no matter what the other players do.

Dutch auction | Auction where prices start high and decrease until a bidder signals willingness to pay, at which point the auction stops.

Dutch auction | Auction where prices start high and decrease until a bidder signals willingness to pay, at which point the auction stops.

economy of scale | Situation that exists when larger scale lowers average cost.

economy of scale | Situation that exists when larger scale lowers average cost.

economy of scope | Situation that exists when producing more related goods lowers average cost.

economy of scope | Situation that exists when producing more related goods lowers average cost.

Edgeworth box | A graphical representation of the exchange problem facing participants in a two-good exchange economy.

Edgeworth box | A graphical representation of the exchange problem facing participants in a two-good exchange economy.

efficient | Maximizing the gains from trade under the assumption that the only people affected by any given transaction are the buyers and the seller.

efficient | Maximizing the gains from trade under the assumption that the only people affected by any given transaction are the buyers and the seller.

elasticity | The percentage change in one variable for a small percentage change in another.

elasticity of demand | The percentage change in one variable for a small percentage change in another.

elasticity of demand | The percentage change in one variable for a small percentage change in another.

elasticity of supply | The percentage increase in quantity supplied resulting from a small percentage increase in price.

elasticity of supply | The percentage increase in quantity supplied resulting from a small percentage increase in price.

Engel curve | Graph that shows the path of consumption as income changes, holding prices constant.

Engel curve | Graph that shows the path of consumption as income changes, holding prices constant.

English auction | Auction where bids increase until no one is willing to top the current bid.

English auction | Auction where bids increase until no one is willing to top the current bid.

equilibrium | Condition that occurs when the pressure for higher prices is balanced by the pressure for lower prices so that the current rate of exchange between buyers and sellers persists.

equilibrium | Condition that occurs when the pressure for higher prices is balanced by the pressure for lower prices so that the current rate of exchange between buyers and sellers persists.

excess burden of taxation | The amount by which the cost of taxation exceeds the taxes raised.

excess burden of taxation | The amount by which the cost of taxation exceeds the taxes raised.

exchange economy | An economy where the supply of each good is just the total endowment of that good, and there is no production.

exchange economy | An economy where the supply of each good is just the total endowment of that good, and there is no production.

external benefits | Benefits obtained by third parties.

external benefits | Benefits obtained by third parties.

external economy of scale | An economy of scale that operates at the industry level, not the individual firm level.

external economy of scale | An economy of scale that operates at the industry level, not the individual firm level.

externality | Any effect on people not involved in a particular transaction.

externality | Any effect on people not involved in a particular transaction.

factor price equalization | Theory that predicts that the value of factors of production should be equalized through trade.

factor price equalization | Theory that predicts that the value of factors of production should be equalized through trade.

Factors of production | Inputs to the production process.

Factors of production | Inputs to the production process.

Federal Trade Commission (FTC) | Federal government agency that enforces the antitrust laws, along with the U.S. Department of Justice (DOJ), and provides consumer protection.

Federal Trade Commission (FTC) | Federal government agency that enforces the antitrust laws, along with the U.S. Department of Justice (DOJ), and provides consumer protection.

first mover advantage | Situation in which choosing first is better than choosing simultaneously.

first mover advantage | Situation in which choosing first is better than choosing simultaneously.

first-order condition | A mathematical condition for optimization stating that the first derivative is zero.

first-order condition | A mathematical condition for optimization stating that the first derivative is zero.

Fixed factors of production | Factors that are not readily moved and thus give particular regions a comparative advantage in the production of some kinds of goods and not in others.

Fixed factors of production | Factors that are not readily moved and thus give particular regions a comparative advantage in the production of some kinds of goods and not in others.

fixed-proportions production function | A production function that requires inputs be used in fixed proportions to produce output.

fixed-proportions production function | A production function that requires inputs be used in fixed proportions to produce output.

flow | The rate of change in the amount of some material that exists from one instant to the next.

flow | The rate of change in the amount of some material that exists from one instant to the next.

Foreclosure | Denying access to necessary inputs.

Foreclosure | Denying access to necessary inputs.

free riders | Individuals who don't contribute to the provision of a public good.

free riders | Individuals who don't contribute to the provision of a public good.

Great Depression | A prolonged and severe economic downturn from 1929 to 1939.

Great Depression | A prolonged and severe economic downturn from 1929 to 1939.

grim trigger strategy | Strategy that involves being nice initially but not nice forever when someone else isn't cooperative.

grim trigger strategy | Strategy that involves being nice initially but not nice forever when someone else isn't cooperative.

gross domestic product (GDP) | The value of traded goods and services produced within the borders of the United States.

gross domestic product (GDP) | The value of traded goods and services produced within the borders of the United States.

hedonic pricing | Method of valuation in which each item is first evaluated separately and then the item values are added together to arrive at a total value.

hedonic pricing | Method of valuation in which each item is first evaluated separately and then the item values are added together to arrive at a total value.

Hirschman-Herfindahl Index (HHI) | The weighted average of the price-cost margins of all firms in the market.

Hirschman-Herfindahl Index (HHI) | The weighted average of the price-cost margins of all firms in the market.

horizontal differentiation | Variety.

horizontal differentiation | Variety.

Hotelling model | Model of imperfect competition in which customers' preferences and products are located by points on the same line segment.

Hotelling model | Model of imperfect competition in which customers' preferences and products are located by points on the same line segment.

households | Groups of people sharing living quarters.

households | Groups of people sharing living quarters.

incidence of the tax | Changes in the price paid for a good based on the amount of tax on the good.

incidence of the tax | Changes in the price paid for a good based on the amount of tax on the good.

income effect | The effect on consumption of a change in income.

income effect | The effect on consumption of a change in income.

indifference curve | The set of goods that produce equal utility; also known as iso-utility curve.

indifference curve | The set of goods that produce equal utility; also known as iso-utility curve.

indirect price discrimination | Offering a menu or set of prices and permitting customers to choose distinct prices.

indirect price discrimination | Offering a menu or set of prices and permitting customers to choose distinct prices.

Individual rationality | Situation in which the outcome offers a present value of profits at least as high as that offered in the worst equilibrium in the stage game from that player's perspective.

Individual rationality | Situation in which the outcome offers a present value of profits at least as high as that offered in the worst equilibrium in the stage game from that player's perspective.

informational asymmetry | Situation in which some parties have information that others do not.

informational asymmetry | Situation in which some parties have information that others do not.

internal rate of return (IRR) | Method of analyzing an investment that solves the equation $NPV = 0$ for the interest rate.

internal rate of return (IRR) | Method of analyzing an investment that solves the equation $NPV = 0$ for the interest rate.

isocost | Line that holds constant the expenditure on inputs.

isocost | Line that holds constant the expenditure on inputs.

isoquant | Curves that describe all the combinations of inputs that produce the same level of output.

isoquant | Curves that describe all the combinations of inputs that produce the same level of output.

iterated elimination of dominated strategies | Eliminating strategies by sequentially removing strategies that are dominated for a player.

iterated elimination of dominated strategies | Eliminating strategies by sequentially removing strategies that are dominated for a player.

law of one price | Identical goods sell at the same price.

law of one price | Identical goods sell at the same price.

linkage principle | The expected price in an auction to sell rises the more the price is linked to the actual value.

linkage principle | The expected price in an auction to sell rises the more the price is linked to the actual value.

local public goods | Goods that are produced and consumed in a limited geographical area.

local public goods | Goods that are produced and consumed in a limited geographical area.

logarithmic scale | Scale on which a straight line gives constant percentage growth.

logarithmic scale | Scale on which a straight line gives constant percentage growth.

long-run average variable cost | The long-run total cost divided by output.

long-run average variable cost | The long-run total cost divided by output.

long-run total cost | The total cost of output with all factors varying.

long-run total cost | The total cost of output with all factors varying.

marginal | Term meaning "the derivative of."

marginal | Term meaning "the derivative of."

marginal product | The derivative of the production function with respect to an input.

marginal product | The derivative of the production function with respect to an input.

marginal rate of substitution (MRS) | The amount extra of one good needed to make up for a decrease in another good, staying on an indifference curve.

marginal rate of substitution (MRS) | The amount extra of one good needed to make up for a decrease in another good, staying on an indifference curve.

marginal rate of technical substitution | The amount of one input needed to make up for a decrease in another input while holding output constant.

marginal rate of technical substitution | The amount of one input needed to make up for a decrease in another input while holding output constant.

marginal revenue | The derivative of revenue pq with respect to quantity.

marginal revenue | The derivative of revenue pq with respect to quantity.

market demand | The quantity purchased by all market participants for each price.

market demand | The quantity purchased by all market participants for each price.

market failure | A situation where a competitive market does not yield the socially efficient outcome.

market failure | A situation where a competitive market does not yield the socially efficient outcome.

market for lemons | A model where sellers are better informed about quality than buyers.

market for lemons | A model where sellers are better informed about quality than buyers.

market power | The ability to increase their price above marginal cost and sustain sales for a long period of time.

market power | The ability to increase their price above marginal cost and sustain sales for a long period of time.

Market supply | The sum of all the individual supply curves for all market participants.

Market supply | The sum of all the individual supply curves for all market participants.

matching pennies | A child's game in which the sum of the payoffs is zero.

matching pennies | A child's game in which the sum of the payoffs is zero.

maximum sustainable yield | Policy that maximizes the long-run average value of a sustainable resource.

maximum sustainable yield | Policy that maximizes the long-run average value of a sustainable resource.

mean variance preferences | Preference that describe people who value risk linearly with the expected return.

mean variance preferences | Preference that describe people who value risk linearly with the expected return.

mechanism | A means of arranging a trade.

mechanism | A means of arranging a trade.

median voter | The voter whose preferences fall in the middle of the range.

median voter | The voter whose preferences fall in the middle of the range.

minimum wage | The minimum amount that a worker can be paid per hour.

minimum wage | The minimum amount that a worker can be paid per hour.

mixed strategies | Random strategies.

mixed strategies | Random strategies.

monopoly | A firm that faces a downward sloping demand and has a choice about what price to charge.

monopoly | A firm that faces a downward sloping demand and has a choice about what price to charge.

Multi-tasking | Performing several activities simultaneously.

Multi-tasking | Performing several activities simultaneously.

Myerson-Satterthwaite theorem | A theorem that shows private information about value may prevent efficient trade.

Myerson-Satterthwaite theorem | A theorem that shows private information about value may prevent efficient trade.

Nash equilibrium | Situation in which a player chooses the strategy that maximizes his or her expected payoff, given the strategies employed by others.

Nash equilibrium | Situation in which a player chooses the strategy that maximizes his or her expected payoff, given the strategies employed by others.

natural monopoly | Situation that arises when a single firm can efficiently serve the entire market because average costs are lower with one firm than with two firms.

natural monopoly | Situation that arises when a single firm can efficiently serve the entire market because average costs are lower with one firm than with two firms.

net present value | The present value of a stream of net payments.

net present value | The present value of a stream of net payments.

network externality | Situation that arises when others' use of a product makes it more valuable to each consumer.

network externality | Situation that arises when others' use of a product makes it more valuable to each consumer.

nonprofit firm | Firm that is prohibited from distributing a profit to its owners.

nonprofit firm | Firm that is prohibited from distributing a profit to its owners.

nonrivalry | Condition in which many people can use a good simultaneously.

nonrivalry | Condition in which many people can use a good simultaneously.

normal good | Goods whose demand increases with income.

normal good | Goods whose demand increases with income.

normative analyses | A study that makes value judgments.

normative analyses | A study that makes value judgments.

numeraire | A good used as a reference good, so that prices are denominated in units of that good.

numeraire | A good used as a reference good, so that prices are denominated in units of that good.

opportunity cost | The value that one forgoes in purchasing a product or undertaking an activity.

opportunity cost | The value that one forgoes in purchasing a product or undertaking an activity.

option | The right to buy or sell at a price determined in advance.

option | The right to buy or sell at a price determined in advance.

Pareto efficiency | Condition that exists when there is no waste or slack in a system, even if it is wildly unfair.

Pareto efficiency | Condition that exists when there is no waste or slack in a system, even if it is wildly unfair.

peak-load pricing | The pricing of a service when demand for it is at its highest.

peak-load pricing | The pricing of a service when demand for it is at its highest.

perfect substitutes | Two goods that can be substituted for each other at a constant rate while maintaining the same output level.

perfect substitutes | Two goods that can be substituted for each other at a constant rate while maintaining the same output level.

permanent income | The present value of the income stream.

permanent income | The present value of the income stream.

Perron-Frobenius theorem | Theorem that states that if **B is a positive matrix (each component positive), then there is an eigenvalue $\lambda > 0$ and an associated positive eigenvector p; and, moreover, λ is the largest (in absolute value) eigenvector of B.**

Perron-Frobenius theorem | Theorem that states that if **B is a positive matrix (each component positive), then there is an eigenvalue $\lambda > 0$ and an associated positive eigenvector p; and, moreover, λ is the largest (in absolute value) eigenvector of B.**

Pigouvian tax | A per-unit tax on a good.

Pigouvian tax | A per-unit tax on a good.

Predatory pricing | Pricing below cost in order to drive a rival out of business.

Predatory pricing | Pricing below cost in order to drive a rival out of business.

price discrimination | Charging distinct customers different prices for the same good.

price discrimination | Charging distinct customers different prices for the same good.

price dispersion | Price variation across stores.

price dispersion | Price variation across stores.

price support | The combination of a minimum price, or price floor, and government purchase of any surplus.

price support | The combination of a minimum price, or price floor, and government purchase of any surplus.

price system | System that involves a specific price for trading *Y* for *X*, and vice versa, that is available to both parties.

price system | System that involves a specific price for trading *Y* for *X*, and vice versa, that is available to both parties.

principal | The entity—person or corporation—on whose behalf an agent works.

principal | The entity—person or corporation—on whose behalf an agent works.

prisoner's dilemma | Game in which the strategies are to confess or not to confess; the first player to confess avoids jail.

prisoner's dilemma | Game in which the strategies are to confess or not to confess; the first player to confess avoids jail.

private benefit | The benefit obtained by the buyer.

private benefit | The benefit obtained by the buyer.

production function | The mapping from inputs to an output or outputs.

production function | The mapping from inputs to an output or outputs.

production possibilities | The collection of feasible outputs of an individual, group or society, or country.

production possibilities | The collection of feasible outputs of an individual, group or society, or country.

production possibilities frontier | The boundary of the production possibilities set.

production possibilities frontier | The boundary of the production possibilities set.

property rights | Rights that come with ownership.

property rights | Rights that come with ownership.

pure strategy Nash equilibrium | A Nash equilibrium in which no player randomizes.

pure strategy Nash equilibrium | A Nash equilibrium in which no player randomizes.

quasilinear utility | Utility that is additively separable.

quasilinear utility | Utility that is additively separable.

quintile | One fifth, or 20%, of a group.

quintile | One fifth, or 20%, of a group.

quota | A maximal production quantity, usually set based on historical production.

quota | A maximal production quantity, usually set based on historical production.

Raising rivals' costs | Increasing input prices as a means of harming a rival.

Raising rivals' costs | Increasing input prices as a means of harming a rival.

Ramsey rule | For resources in fixed supply, prices rise at the interest rate.

Ramsey rule | For resources in fixed supply, prices rise at the interest rate.

rate of return regulation | Price-limiting strategy that limits the profitability of firms.

rate of return regulation | Price-limiting strategy that limits the profitability of firms.

recessions | Periods marked by a drop in gross domestic output.

recessions | Periods marked by a drop in gross domestic output.

relevant antitrust market | A set of products sold in a geographic area in which a hypothetical monopoly can profitably raise price.

relevant antitrust market | A set of products sold in a geographic area in which a hypothetical monopoly can profitably raise price.

reservation price | The maximum acceptable price a consumer will pay without visiting another supplier.

reservation price | The maximum acceptable price a consumer will pay without visiting another supplier.

Reserve prices | Minimum bids in an auction.

Reserve prices | Minimum bids in an auction.

revealed preference | Kind that states that choice implies preference.

revealed preference | Kind that states that choice implies preference.

revenue equivalence | Situation in which two auctions produce the same price on average.

revenue equivalence | Situation in which two auctions produce the same price on average.

Ricardian theory | Theory that suggests nations, responding to price incentives, will specialize in the production of goods in which they have a comparative advantage, and purchase the goods in which they have a comparative disadvantage.

Ricardian theory | Theory that suggests nations, responding to price incentives, will specialize in the production of goods in which they have a comparative advantage, and purchase the goods in which they have a comparative disadvantage.

risk premium | The difference between the expected payoff and the certainty equivalent.

risk premium | The difference between the expected payoff and the certainty equivalent.

risk premium | The difference between the expected payoff and the certainty equivalent.

risk premium | The difference between the expected payoff and the certainty equivalent.

risk-averse | Preferring the expected value of a gamble to the gamble.

risk-averse | Preferring the expected value of a gamble to the gamble.

Robinson-Patman | Section of the Clayton Act prohibiting price discrimination that lessens competition.

Robinson-Patman | Section of the Clayton Act prohibiting price discrimination that lessens competition.

satiation | The point at which increased consumption does not increase utility.

satiation | The point at which increased consumption does not increase utility.

sealed-bid auction | Auction where bidders simultaneously submit sealed bids, and the highest bidder wins and pays the highest bid.

sealed-bid auction | Auction where bidders simultaneously submit sealed bids, and the highest bidder wins and pays the highest bid.

second welfare theorem of general equilibrium | Theorem that states that any Pareto-efficient point is an equilibrium of the price system for some endowment.

second welfare theorem of general equilibrium | Theorem that states that any Pareto-efficient point is an equilibrium of the price system for some endowment.

second-order condition | A mathematical condition for maximization stating that the second derivative is nonpositive.

second-order condition | A mathematical condition for maximization stating that the second derivative is nonpositive.

self-interested behavior | Selfishness.

self-interested behavior | Selfishness.

shadow value | The value associated with a constraint.

shadow value | The value associated with a constraint.

sharing rule | In agency theory, sharing a fixed proportion of the output with the agent.

sharing rule | In agency theory, sharing a fixed proportion of the output with the agent.

Sherman Act | The first U.S. antitrust law; it makes restraint of trade (monopolization) illegal.

Sherman Act | The first U.S. antitrust law; it makes restraint of trade (monopolization) illegal.

short run | Time interval in which not all inputs can be changed.

short run | Time interval in which not all inputs can be changed.

short-run average variable cost | The average cost ignoring the investment in capital equipment.

short-run average variable cost | The average cost ignoring the investment in capital equipment.

short-run equilibrium | The point where short-run demand equals short-run supply.

short-run equilibrium | The point where short-run demand equals short-run supply.

short-run fixed cost | The difference between variable cost and total cost.

short-run fixed cost | The difference between variable cost and total cost.

short-run marginal cost | The derivative of short-run total cost with respect to output.

short-run marginal cost | The derivative of short-run total cost with respect to output.

short-run total cost | The total cost of output with only short-run factors varying.

short-run total cost | The total cost of output with only short-run factors varying.

short-run variable cost | The total cost minus the cost of producing zero units.

short-run variable cost | The total cost minus the cost of producing zero units.

shortage | Condition in which the quantity demanded exceeds the quantity supplied.

shortage | Condition in which the quantity demanded exceeds the quantity supplied.

signaling | Expenditures of time or money whose purpose is to convince others of something.

signaling | Expenditures of time or money whose purpose is to convince others of something.

social costs | The total costs for all parties.

social costs | The total costs for all parties.

stage game | The game that is repeated in a supergame.

stage game | The game that is repeated in a supergame.

steady state | Condition in which the variables stay constant and forces are balanced.

steady state | Condition in which the variables stay constant and forces are balanced.

strategic equivalence | Situation in which two games are strategically equivalent if they have the same strategies (after a renaming) and the strategies lead to the same outcomes.

strategic equivalence | Situation in which two games are strategically equivalent if they have the same strategies (after a renaming) and the strategies lead to the same outcomes.

strike price | The amount one pays to obtain a share of stock.

strike price | The amount one pays to obtain a share of stock.

subgame perfection | The idea that every stage of a dynamic game is played in an optimal way.

subgame perfection | The idea that every stage of a dynamic game is played in an optimal way.

substitute | For a given good x, a good whose consumption decreases the value of x.

substitute | For a given good x, a good whose consumption decreases the value of x.

substitute in supply | A good whose cost rises as the amount produced of another good rises.

substitute in supply | A good whose cost rises as the amount produced of another good rises.

substitution effect | The effect on consumption of a change in the relative price, with a sufficient change in income to keep the consumer on the same utility isoquant.

substitution effect | The effect on consumption of a change in the relative price, with a sufficient change in income to keep the consumer on the same utility isoquant.

sunk cost fallacy | A psychological tendency to invest more once one has made a significant nonrecoverable investment, even when subsequent investment isn't warranted.

sunk cost fallacy | A psychological tendency to invest more once one has made a significant nonrecoverable investment, even when subsequent investment isn't warranted.

supply | The function that gives the quantity offered for sale as a function of price.

supply | The function that gives the quantity offered for sale as a function of price.

support price | The minimum price the government is willing to pay for a product.

support price | The minimum price the government is willing to pay for a product.

surplus | Condition in which the quantity supplied exceeds the quantity demanded.

surplus | Condition in which the quantity supplied exceeds the quantity demanded.

the battle of the sexes | A coordination game that involves a married couple who are going to meet each other after work but haven't decided where they are meeting.

the battle of the sexes | A coordination game that involves a married couple who are going to meet each other after work but haven't decided where they are meeting.

total expenditure | Price times the quantity purchased.

total expenditure | Price times the quantity purchased.

Tradable permits | Quotas for pollution that can be exchanged to create a market in the right to pollute.

Tradable permits | Quotas for pollution that can be exchanged to create a market in the right to pollute.

tragedy of the commons | Overuse of a resource due to lack of ownership.

tragedy of the commons | Overuse of a resource due to lack of ownership.

Transfers to individuals | Direct payments to individuals in the form of a check.

Transfers to individuals | Direct payments to individuals in the form of a check.

Treasury bill | A bond that has a fixed redemption value.

Treasury bill | A bond that has a fixed redemption value.

two-part pricing | Price discrimination featuring a fixed charge plus a marginal charge.

two-part pricing | Price discrimination featuring a fixed charge plus a marginal charge.

two-part test | In antitrust, the requirement for an antitrust violation that a firm had monopoly power and abused that power.

two-part test | In antitrust, the requirement for an antitrust violation that a firm had monopoly power and abused that power.

U.S. Department of Justice | Federal government agency that enforces the antitrust laws, along with the Federal Trade Commission (FTC), among other responsibilities.

U.S. Department of Justice | Federal government agency that enforces the antitrust laws, along with the Federal Trade Commission (FTC), among other responsibilities.

unilateral effect | In antitrust, a situation in which merging parties increase price unilaterally after the merger is consummated.

unilateral effect | In antitrust, a situation in which merging parties increase price unilaterally after the merger is consummated.

unitary elasticity | When elasticity is equal to one.

unitary elasticity | When elasticity is equal to one.

usury laws | Laws that limit the legal interest rate that can be charged.

usury laws | Laws that limit the legal interest rate that can be charged.

Utility | Flow of pleasure or happiness that a person enjoys.

Utility | Flow of pleasure or happiness that a person enjoys.

value of the marginal product | The marginal product times the price of the output.

value of the marginal product | The marginal product times the price of the output.

vertical merger | A merger between an input supplier and an input buyer.

vertical merger | A merger between an input supplier and an input buyer.

Vickrey auction | Auction where bidders bid simultaneously, and the highest bidder wins and pays the second-highest bid.

Vickrey auction | Auction where bidders bid simultaneously, and the highest bidder wins and pays the second-highest bid.

von Neumann-Morgenstern utility function | The value of each outcome, constructed in such a way that a consumer values gambles as if they were the expected utility.

von Neumann-Morgenstern utility function | The value of each outcome, constructed in such a way that a consumer values gambles as if they were the expected utility.

Welfare analysis | A normative analysis that trades off gains and losses to different individuals.

Welfare analysis | A normative analysis that trades off gains and losses to different individuals.

willingness-to-pay | The amount a customer is willing and able to pay for a good.

willingness-to-pay | The amount a customer is willing and able to pay for a good.

winner's curse | The bidder who most overestimates the value of the object wins the bidding.

winner's curse | The bidder who most overestimates the value of the object wins the bidding.

Detailed Licensing Detailed LicensingEdit Update titlecancel

This page is a draft and is under active development.

Overview

Title: Introduction to Economic Analysis (LibreTexts)

Webpages: 137

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 3.0: 96.4% (132 pages)
- Undeclared: 3.6% (5 pages)

By Page

- Introduction to Economic Analysis (LibreTexts) - *CC BY-NC-SA 3.0*
  - Front Matter - *CC BY-NC-SA 3.0*
    - TitlePage - *CC BY-NC-SA 3.0*
    - InfoPage - *CC BY-NC-SA 3.0*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: What Is Economics? - *CC BY-NC-SA 3.0*
    - 1.1: Normative and Positive Theories - *CC BY-NC-SA 3.0*
    - 1.2: Opportunity Cost - *CC BY-NC-SA 3.0*
    - 1.3: Economic Reasoning and Analysis - *CC BY-NC-SA 3.0*
  - 2: Supply and Demand - *CC BY-NC-SA 3.0*
    - 2.1: Demand and Consumer Surplus - *CC BY-NC-SA 3.0*
    - 2.2: Supply and Profit - *CC BY-NC-SA 3.0*
    - 2.3: Market Demand and Supply - *CC BY-NC-SA 3.0*
    - 2.4: Equilibrium - *CC BY-NC-SA 3.0*
    - 2.5: Changes in Demand and Supply - *CC BY-NC-SA 3.0*
  - 3: Quantification - *CC BY-NC-SA 3.0*
    - 3.1: Elasticity - *CC BY-NC-SA 3.0*
    - 3.2: Supply and Demand Changes - *CC BY-NC-SA 3.0*
  - 4: The U.S. Economy - *CC BY-NC-SA 3.0*
    - 4.1: Basic Demographics - *CC BY-NC-SA 3.0*
    - 4.2: Education - *CC BY-NC-SA 3.0*
    - 4.3: Households and Consumption - *CC BY-NC-SA 3.0*
    - 4.4: Production - *CC BY-NC-SA 3.0*
    - 4.5: Government - *CC BY-NC-SA 3.0*
    - 4.6: Trade - *CC BY-NC-SA 3.0*
    - 4.7: Fluctuations - *CC BY-NC-SA 3.0*
  - 5: Government Interventions - *CC BY-NC-SA 3.0*
    - 5.1: Effects of Taxes - *CC BY-NC-SA 3.0*
    - 5.2: Incidence of Taxes - *CC BY-NC-SA 3.0*
    - 5.3: Excess Burden of Taxation - *CC BY-NC-SA 3.0*
    - 5.4: Price Floors and Ceilings - *CC BY-NC-SA 3.0*
    - 5.5: The Politics of Price Controls - *CC BY-NC-SA 3.0*
    - 5.6: Price Supports - *CC BY-NC-SA 3.0*
    - 5.7: Quantity Restrictions and Quotas - *CC BY-NC-SA 3.0*
  - 6: Trade - *CC BY-NC-SA 3.0*
    - 6.1: Production Possibilities Frontier - *CC BY-NC-SA 3.0*
    - 6.2: Comparative and Absolute Advantage - *CC BY-NC-SA 3.0*
    - 6.3: Factors of Production - *CC BY-NC-SA 3.0*
    - 6.4: International Trade - *CC BY-NC-SA 3.0*
  - 7: Externalities - *CC BY-NC-SA 3.0*
    - 7.1: External Effects - *CC BY-NC-SA 3.0*
    - 7.2: Pigouvian Taxes - *CC BY-NC-SA 3.0*
    - 7.3: Quotas - *CC BY-NC-SA 3.0*
    - 7.4: Tradable Permits and Auctions - *CC BY-NC-SA 3.0*
    - 7.5: Coasian Bargaining - *CC BY-NC-SA 3.0*
    - 7.6: Fishing and Extinction - *CC BY-NC-SA 3.0*
  - 8: Public Goods - *CC BY-NC-SA 3.0*
    - 8.1: Free Riders - *CC BY-NC-SA 3.0*
    - 8.2: Provision With Taxation - *CC BY-NC-SA 3.0*
    - 8.3: Local Public Goods - *CC BY-NC-SA 3.0*
  - 9: Producer Theory- Costs - *CC BY-NC-SA 3.0*
    - 9.1: Types of Firms - *CC BY-NC-SA 3.0*
    - 9.2: Production Functions - *CC BY-NC-SA 3.0*
    - 9.3: Profit Maximization - *CC BY-NC-SA 3.0*
    - 9.4: The Shadow Value - *CC BY-NC-SA 3.0*

- 9.5: Input Demand - *CC BY-NC-SA 3.0*
- 9.6: Myriad Costs - *CC BY-NC-SA 3.0*

- 10: Producer Theory- Dynamics - *CC BY-NC-SA 3.0*
  - 10.1: Reactions of Competitive Firms - *CC BY-NC-SA 3.0*
  - 10.2: Economies of Scale and Scope - *CC BY-NC-SA 3.0*
  - 10.3: Dynamics With Constant Costs - *CC BY-NC-SA 3.0*
  - 10.4: General Long-run Dynamics - *CC BY-NC-SA 3.0*
- 11: Investment - *CC BY-NC-SA 3.0*
  - 11.1: Present Value - *CC BY-NC-SA 3.0*
  - 11.2: Investment - *CC BY-NC-SA 3.0*
  - 11.3: Investment Under Uncertainty - *CC BY-NC-SA 3.0*
  - 11.4: Resource Extraction - *CC BY-NC-SA 3.0*
  - 11.5: A Time to Harvest - *CC BY-NC-SA 3.0*
  - 11.6: Collectibles - *CC BY-NC-SA 3.0*
  - 11.7: Summer Wheat - *CC BY-NC-SA 3.0*
- 12: Consumer Theory - *CC BY-NC-SA 3.0*
  - 12.1: Utility Maximization - *CC BY-NC-SA 3.0*
  - 12.2: Budget or Feasible Set - *CC BY-NC-SA 3.0*
  - 12.3: Isoquants - *CC BY-NC-SA 3.0*
  - 12.4: Examples - *CC BY-NC-SA 3.0*
  - 12.5: Substitution Effects - *CC BY-NC-SA 3.0*
  - 12.6: Income Effects - *CC BY-NC-SA 3.0*
  - 12.7: Mathematical Cleanup - *CC BY-NC-SA 3.0*
- 13: Applied Consumer Theory - *CC BY-NC-SA 3.0*
  - 13.1: Labor Supply - *CC BY-NC-SA 3.0*
  - 13.2: Urban Real Estate Prices - *CC BY-NC-SA 3.0*
  - 13.3: Dynamic Choice - *CC BY-NC-SA 3.0*
  - 13.4: Risk Aversion - *CC BY-NC-SA 3.0*
  - 13.5: Search - *CC BY-NC-SA 3.0*
- 14: General Equilibrium - *CC BY-NC-SA 3.0*
  - 14.1: Edgeworth Box - *CC BY-NC-SA 3.0*
  - 14.2: Equilibrium With Price System - *CC BY-NC-SA 3.0*
  - 14.3: General Equilibrium - *CC BY-NC-SA 3.0*
- 15: Monopoly - *CC BY-NC-SA 3.0*
  - 15.1: Sources of Monopoly - *CC BY-NC-SA 3.0*
  - 15.2: Basic Analysis - *CC BY-NC-SA 3.0*
  - 15.3: Effect of Taxes - *CC BY-NC-SA 3.0*
  - 15.4: Price Discrimination - *CC BY-NC-SA 3.0*
  - 15.5: Welfare Effects - *CC BY-NC-SA 3.0*
  - 15.6: Natural Monopoly - *CC BY-NC-SA 3.0*
  - 15.7: Peak-load Pricing - *CC BY-NC-SA 3.0*
- 16: Games and Strategic Behavior - *CC BY-NC-SA 3.0*
  - 16.1: Matrix Games - *CC BY-NC-SA 3.0*
  - 16.2: Nash Equilibrium - *CC BY-NC-SA 3.0*
  - 16.3: Mixed Strategies - *CC BY-NC-SA 3.0*
  - 16.4: Examples - *CC BY-NC-SA 3.0*
  - 16.5: Subgame Perfection - *CC BY-NC-SA 3.0*
  - 16.6: Supergames - *CC BY-NC-SA 3.0*
- 17: Imperfect Competition - *CC BY-NC-SA 3.0*
  - 17.1: Cournot Oligopoly - *CC BY-NC-SA 3.0*
  - 17.2: Cournot Industry Performance - *CC BY-NC-SA 3.0*
  - 17.3: Hotelling Differentiation - *CC BY-NC-SA 3.0*
  - 17.4: The Circle Model - *CC BY-NC-SA 3.0*
- 18: Information - *CC BY-NC-SA 3.0*
  - 18.1: Market for Lemons - *CC BY-NC-SA 3.0*
  - 18.2: Myerson-Satterthwaite Theorem - *CC BY-NC-SA 3.0*
  - 18.3: Signaling - *CC BY-NC-SA 3.0*
  - 18.4: Search and Price Dispersion - *CC BY-NC-SA 3.0*
- 19: Agency Theory - *CC BY-NC-SA 3.0*
  - 19.1: Principals and Agents - *CC BY-NC-SA 3.0*
  - 19.2: Cost of Providing Incentives - *CC BY-NC-SA 3.0*
  - 19.3: Multi-tasking without Homogeneity - *CC BY-NC-SA 3.0*
  - 19.4: Multi-tasking without Homogeneity - *CC BY-NC-SA 3.0*
- 20: Auctions - *CC BY-NC-SA 3.0*
  - 20.1: English Auction - *CC BY-NC-SA 3.0*
  - 20.2: Sealed-bid Auction - *CC BY-NC-SA 3.0*
  - 20.3: Dutch Auction - *CC BY-NC-SA 3.0*
  - 20.4: Vickrey Auction - *CC BY-NC-SA 3.0*
  - 20.5: The Winner's Curse and Linkage - *CC BY-NC-SA 3.0*
  - 20.6: Auction Design - *CC BY-NC-SA 3.0*
- 21: Antitrust - *CC BY-NC-SA 3.0*
  - 21.1: Sherman Act - *CC BY-NC-SA 3.0*
  - 21.2: Clayton Act - *CC BY-NC-SA 3.0*
  - 21.3: Price Fixing - *CC BY-NC-SA 3.0*
  - 21.4: Mergers - *CC BY-NC-SA 3.0*
- Back Matter - *CC BY-NC-SA 3.0*
  - Index - *CC BY-NC-SA 3.0*
  - Index - *Undeclared*
  - Glossary - *CC BY-NC-SA 3.0*
  - Detailed Licensing - *Undeclared*
  - Detailed Licensing - *Undeclared*

Detailed Licensing Detailed LicensingEdit Update titlecancel

Overview

Title: Introduction to Economic Analysis (LibreTexts)

Webpages: 137

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 3.0: 96.4% (132 pages)
- Undeclared: 3.6% (5 pages)

By Page

- Introduction to Economic Analysis (LibreTexts) - *CC BY-NC-SA 3.0*
  - Front Matter - *CC BY-NC-SA 3.0*
    - TitlePage - *CC BY-NC-SA 3.0*
    - InfoPage - *CC BY-NC-SA 3.0*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: What Is Economics? - *CC BY-NC-SA 3.0*
    - 1.1: Normative and Positive Theories - *CC BY-NC-SA 3.0*
    - 1.2: Opportunity Cost - *CC BY-NC-SA 3.0*
    - 1.3: Economic Reasoning and Analysis - *CC BY-NC-SA 3.0*
  - 2: Supply and Demand - *CC BY-NC-SA 3.0*
    - 2.1: Demand and Consumer Surplus - *CC BY-NC-SA 3.0*
    - 2.2: Supply and Profit - *CC BY-NC-SA 3.0*
    - 2.3: Market Demand and Supply - *CC BY-NC-SA 3.0*
    - 2.4: Equilibrium - *CC BY-NC-SA 3.0*
    - 2.5: Changes in Demand and Supply - *CC BY-NC-SA 3.0*
  - 3: Quantification - *CC BY-NC-SA 3.0*
    - 3.1: Elasticity - *CC BY-NC-SA 3.0*
    - 3.2: Supply and Demand Changes - *CC BY-NC-SA 3.0*
  - 4: The U.S. Economy - *CC BY-NC-SA 3.0*
    - 4.1: Basic Demographics - *CC BY-NC-SA 3.0*
    - 4.2: Education - *CC BY-NC-SA 3.0*
    - 4.3: Households and Consumption - *CC BY-NC-SA 3.0*
    - 4.4: Production - *CC BY-NC-SA 3.0*
    - 4.5: Government - *CC BY-NC-SA 3.0*
    - 4.6: Trade - *CC BY-NC-SA 3.0*
    - 4.7: Fluctuations - *CC BY-NC-SA 3.0*
  - 5: Government Interventions - *CC BY-NC-SA 3.0*
    - 5.1: Effects of Taxes - *CC BY-NC-SA 3.0*
    - 5.2: Incidence of Taxes - *CC BY-NC-SA 3.0*
    - 5.3: Excess Burden of Taxation - *CC BY-NC-SA 3.0*
    - 5.4: Price Floors and Ceilings - *CC BY-NC-SA 3.0*
    - 5.5: The Politics of Price Controls - *CC BY-NC-SA 3.0*
    - 5.6: Price Supports - *CC BY-NC-SA 3.0*
    - 5.7: Quantity Restrictions and Quotas - *CC BY-NC-SA 3.0*
  - 6: Trade - *CC BY-NC-SA 3.0*
    - 6.1: Production Possibilities Frontier - *CC BY-NC-SA 3.0*
    - 6.2: Comparative and Absolute Advantage - *CC BY-NC-SA 3.0*
    - 6.3: Factors of Production - *CC BY-NC-SA 3.0*
    - 6.4: International Trade - *CC BY-NC-SA 3.0*
  - 7: Externalities - *CC BY-NC-SA 3.0*
    - 7.1: External Effects - *CC BY-NC-SA 3.0*
    - 7.2: Pigouvian Taxes - *CC BY-NC-SA 3.0*
    - 7.3: Quotas - *CC BY-NC-SA 3.0*
    - 7.4: Tradable Permits and Auctions - *CC BY-NC-SA 3.0*
    - 7.5: Coasian Bargaining - *CC BY-NC-SA 3.0*
    - 7.6: Fishing and Extinction - *CC BY-NC-SA 3.0*
  - 8: Public Goods - *CC BY-NC-SA 3.0*
    - 8.1: Free Riders - *CC BY-NC-SA 3.0*
    - 8.2: Provision With Taxation - *CC BY-NC-SA 3.0*
    - 8.3: Local Public Goods - *CC BY-NC-SA 3.0*
  - 9: Producer Theory- Costs - *CC BY-NC-SA 3.0*
    - 9.1: Types of Firms - *CC BY-NC-SA 3.0*
    - 9.2: Production Functions - *CC BY-NC-SA 3.0*
    - 9.3: Profit Maximization - *CC BY-NC-SA 3.0*
    - 9.4: The Shadow Value - *CC BY-NC-SA 3.0*
    - 9.5: Input Demand - *CC BY-NC-SA 3.0*
    - 9.6: Myriad Costs - *CC BY-NC-SA 3.0*
  - 10: Producer Theory- Dynamics - *CC BY-NC-SA 3.0*

- 10.1: Reactions of Competitive Firms - *CC BY-NC-SA 3.0*
- 10.2: Economies of Scale and Scope - *CC BY-NC-SA 3.0*
- 10.3: Dynamics With Constant Costs - *CC BY-NC-SA 3.0*
- 10.4: General Long-run Dynamics - *CC BY-NC-SA 3.0*

11: Investment - *CC BY-NC-SA 3.0*
- 11.1: Present Value - *CC BY-NC-SA 3.0*
- 11.2: Investment - *CC BY-NC-SA 3.0*
- 11.3: Investment Under Uncertainty - *CC BY-NC-SA 3.0*
- 11.4: Resource Extraction - *CC BY-NC-SA 3.0*
- 11.5: A Time to Harvest - *CC BY-NC-SA 3.0*
- 11.6: Collectibles - *CC BY-NC-SA 3.0*
- 11.7: Summer Wheat - *CC BY-NC-SA 3.0*

12: Consumer Theory - *CC BY-NC-SA 3.0*
- 12.1: Utility Maximization - *CC BY-NC-SA 3.0*
- 12.2: Budget or Feasible Set - *CC BY-NC-SA 3.0*
- 12.3: Isoquants - *CC BY-NC-SA 3.0*
- 12.4: Examples - *CC BY-NC-SA 3.0*
- 12.5: Substitution Effects - *CC BY-NC-SA 3.0*
- 12.6: Income Effects - *CC BY-NC-SA 3.0*
- 12.7: Mathematical Cleanup - *CC BY-NC-SA 3.0*

13: Applied Consumer Theory - *CC BY-NC-SA 3.0*
- 13.1: Labor Supply - *CC BY-NC-SA 3.0*
- 13.2: Urban Real Estate Prices - *CC BY-NC-SA 3.0*
- 13.3: Dynamic Choice - *CC BY-NC-SA 3.0*
- 13.4: Risk Aversion - *CC BY-NC-SA 3.0*
- 13.5: Search - *CC BY-NC-SA 3.0*

14: General Equilibrium - *CC BY-NC-SA 3.0*
- 14.1: Edgeworth Box - *CC BY-NC-SA 3.0*
- 14.2: Equilibrium With Price System - *CC BY-NC-SA 3.0*
- 14.3: General Equilibrium - *CC BY-NC-SA 3.0*

15: Monopoly - *CC BY-NC-SA 3.0*
- 15.1: Sources of Monopoly - *CC BY-NC-SA 3.0*
- 15.2: Basic Analysis - *CC BY-NC-SA 3.0*
- 15.3: Effect of Taxes - *CC BY-NC-SA 3.0*
- 15.4: Price Discrimination - *CC BY-NC-SA 3.0*
- 15.5: Welfare Effects - *CC BY-NC-SA 3.0*
- 15.6: Natural Monopoly - *CC BY-NC-SA 3.0*
- 15.7: Peak-load Pricing - *CC BY-NC-SA 3.0*

16: Games and Strategic Behavior - *CC BY-NC-SA 3.0*
- 16.1: Matrix Games - *CC BY-NC-SA 3.0*
- 16.2: Nash Equilibrium - *CC BY-NC-SA 3.0*
- 16.3: Mixed Strategies - *CC BY-NC-SA 3.0*
- 16.4: Examples - *CC BY-NC-SA 3.0*
- 16.5: Subgame Perfection - *CC BY-NC-SA 3.0*
- 16.6: Supergames - *CC BY-NC-SA 3.0*

17: Imperfect Competition - *CC BY-NC-SA 3.0*
- 17.1: Cournot Oligopoly - *CC BY-NC-SA 3.0*
- 17.2: Cournot Industry Performance - *CC BY-NC-SA 3.0*
- 17.3: Hotelling Differentiation - *CC BY-NC-SA 3.0*
- 17.4: The Circle Model - *CC BY-NC-SA 3.0*

18: Information - *CC BY-NC-SA 3.0*
- 18.1: Market for Lemons - *CC BY-NC-SA 3.0*
- 18.2: Myerson-Satterthwaite Theorem - *CC BY-NC-SA 3.0*
- 18.3: Signaling - *CC BY-NC-SA 3.0*
- 18.4: Search and Price Dispersion - *CC BY-NC-SA 3.0*

19: Agency Theory - *CC BY-NC-SA 3.0*
- 19.1: Principals and Agents - *CC BY-NC-SA 3.0*
- 19.2: Cost of Providing Incentives - *CC BY-NC-SA 3.0*
- 19.3: Multi-tasking without Homogeneity - *CC BY-NC-SA 3.0*
- 19.4: Multi-tasking without Homogeneity - *CC BY-NC-SA 3.0*

20: Auctions - *CC BY-NC-SA 3.0*
- 20.1: English Auction - *CC BY-NC-SA 3.0*
- 20.2: Sealed-bid Auction - *CC BY-NC-SA 3.0*
- 20.3: Dutch Auction - *CC BY-NC-SA 3.0*
- 20.4: Vickrey Auction - *CC BY-NC-SA 3.0*
- 20.5: The Winner's Curse and Linkage - *CC BY-NC-SA 3.0*
- 20.6: Auction Design - *CC BY-NC-SA 3.0*

21: Antitrust - *CC BY-NC-SA 3.0*
- 21.1: Sherman Act - *CC BY-NC-SA 3.0*
- 21.2: Clayton Act - *CC BY-NC-SA 3.0*
- 21.3: Price Fixing - *CC BY-NC-SA 3.0*
- 21.4: Mergers - *CC BY-NC-SA 3.0*

Back Matter - *CC BY-NC-SA 3.0*
- Index - *CC BY-NC-SA 3.0*
- Index - *Undeclared*
- Glossary - *CC BY-NC-SA 3.0*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*