

READING

13

Demand and Supply Analysis: Introduction

by Richard V. Eastin, PhD, and Gary L. Arbogast, CFA

Richard V. Eastin, PhD, is at the University of Southern California (USA). Gary L. Arbogast, CFA (USA).

LEARNING OUTCOMES

Mastery	<i>The candidate should be able to:</i>
<input type="checkbox"/>	a. distinguish among types of markets;
<input type="checkbox"/>	b. explain the principles of demand and supply;
<input type="checkbox"/>	c. describe causes of shifts in and movements along demand and supply curves;
<input type="checkbox"/>	d. describe the process of aggregating demand and supply curves;
<input type="checkbox"/>	e. describe the concept of equilibrium (partial and general), and mechanisms by which markets achieve equilibrium;
<input type="checkbox"/>	f. distinguish between stable and unstable equilibria, including price bubbles, and identify instances of such equilibria;
<input type="checkbox"/>	g. calculate and interpret individual and aggregate demand, and inverse demand and supply functions, and interpret individual and aggregate demand and supply curves;
<input type="checkbox"/>	h. calculate and interpret the amount of excess demand or excess supply associated with a non-equilibrium price;
<input type="checkbox"/>	i. describe types of auctions and calculate the winning price(s) of an auction;
<input type="checkbox"/>	j. calculate and interpret consumer surplus, producer surplus, and total surplus;
<input type="checkbox"/>	k. describe how government regulation and intervention affect demand and supply;
<input type="checkbox"/>	l. forecast the effect of the introduction and the removal of a market interference (e.g., a price floor or ceiling) on price and quantity;
<input type="checkbox"/>	m. calculate and interpret price, income, and cross-price elasticities of demand and describe factors that affect each measure.

1

INTRODUCTION

In a general sense, **economics** is the study of production, distribution, and consumption and can be divided into two broad areas of study: macroeconomics and microeconomics. **Macroeconomics** deals with aggregate economic quantities, such as national output and national income. Macroeconomics has its roots in **microeconomics**, which deals with markets and decision making of individual economic units, including consumers and businesses. Microeconomics is a logical starting point for the study of economics.

This reading focuses on a fundamental subject in microeconomics: demand and supply analysis. **Demand and supply analysis** is the study of how buyers and sellers interact to determine transaction prices and quantities. As we will see, prices simultaneously reflect both the value to the buyer of the next (or marginal) unit and the cost to the seller of that unit. In private enterprise market economies, which are the chief concern of investment analysts, demand and supply analysis encompasses the most basic set of microeconomic tools.

Traditionally, microeconomics classifies private economic units into two groups: consumers (or households) and firms. These two groups give rise, respectively, to the theory of the consumer and theory of the firm as two branches of study. The **theory of the consumer** deals with **consumption** (the demand for goods and services) by utility-maximizing individuals (i.e., individuals who make decisions that maximize the satisfaction received from present and future consumption). The **theory of the firm** deals with the supply of goods and services by profit-maximizing firms. The theory of the consumer and the theory of the firm are important because they help us understand the foundations of demand and supply. Subsequent readings will focus on the theory of the consumer and the theory of the firm.

Investment analysts, particularly equity and credit analysts, must regularly analyze products and services, their costs, prices, possible substitutes, and complements, to reach conclusions about a company's profitability and business risk (risk relating to operating profits). Furthermore, unless the analyst has a sound understanding of the demand and supply model of markets, he or she cannot hope to forecast how external events—such as a shift in consumer tastes or changes in taxes and subsidies or other intervention in markets—will influence a firm's revenue, earnings, and cash flows.

Having grasped the tools and concepts presented in this reading, the reader should also be able to understand many important economic relations and facts and be able to answer questions, such as:

- Why do consumers usually buy more when the price falls? Is it irrational to violate this “law of demand”?
- What are appropriate measures of how sensitive the quantity demanded or supplied is to changes in price, income, and prices of other goods? What affects those sensitivities?
- If a firm lowers its price, will its total revenue also fall? Are there conditions under which revenue might rise as price falls and what are those? Why?
- What is an appropriate measure of the total value consumers or producers receive from the opportunity to buy and sell goods and services in a free market? How might government intervention reduce that value, and what is an appropriate measure of that loss?
- What tools are available that help us frame the trade-offs that consumers and investors face as they must give up one opportunity to pursue another?

- Is it reasonable to expect markets to converge to an equilibrium price? What are the conditions that would make that equilibrium stable or unstable in response to external shocks?
- How do different types of auctions affect price discovery?

This reading is organized as follows. Section 2 explains how economists classify markets. Section 3 covers the basic principles and concepts of demand and supply analysis of markets. Section 4 introduces measures of sensitivity of demand to changes in prices and income. A summary and practice problems conclude the reading.

TYPES OF MARKETS

2

Analysts must understand the demand and supply model of markets because all firms buy and sell in markets. Investment analysts need at least a basic understanding of those markets and the demand and supply model that provides a framework for analyzing them.

Markets are broadly classified as factor markets or goods markets. **Factor markets** are markets for the purchase and sale of factors of production. In capitalist private enterprise economies, households own the factors of production (the land, labor, physical capital, and materials used in production). **Goods markets** are markets for the output of production. From an economics perspective, firms, which ultimately are owned by individuals either singly or in some corporate form, are organizations that buy the services of those factors. Firms then transform those services into intermediate or final goods and services. (**Intermediate goods and services** are those purchased for use as inputs to produce other goods and services, whereas final goods and services are in the final form purchased by households.) These two types of interaction between the household sector and the firm sector—those related to goods and those related to services—take place in factor markets and goods markets, respectively.

In the factor market for labor, households are sellers and firms are buyers. In goods markets: firms are sellers and both households and firms are buyers. For example, firms are buyers of capital goods (such as equipment) and intermediate goods, while households are buyers of a variety of durable and non-durable goods. Generally, market interactions are *voluntary*. Firms offer their products for sale when they believe the payment they will receive exceeds their cost of production. Households are willing to purchase goods and services when the value they expect to receive from them exceeds the payment necessary to acquire them. Whenever the perceived value of a good exceeds the expected cost to produce it, a potential trade can take place. This fact may seem obvious, but it is fundamental to our understanding of markets. If a buyer values something more than a seller, not only is there an opportunity for an exchange, but that exchange will make *both* parties better off.

In one type of factor market, called **labor markets**, households offer to sell their labor services when the payment they expect to receive exceeds the value of the leisure time they must forgo. In contrast, firms hire workers when they judge that the value of the productivity of workers is greater than the cost of employing them. A major source of household income and a major cost to firms is compensation paid in exchange for labor services.

Additionally, households typically choose to spend less on consumption than they earn from their labor. This behavior is called **saving**, through which households can accumulate financial capital, the returns on which can produce other sources of household income, such as interest, dividends, and capital gains. Households may choose to lend their accumulated savings (in exchange for interest) or invest it in ownership claims

in firms (in hopes of receiving dividends and capital gains). Households make these savings choices when their anticipated future returns are judged to be more valuable today than the present consumption that households must sacrifice when they save.

Indeed, a major purpose of financial institutions and markets is to enable the transfer of these savings into capital investments. Firms use **capital markets** (markets for long-term financial capital—that is, markets for long-term claims on firms' assets and cash flows) to sell debt (in bond markets) or equity (in equity markets) in order to raise funds to invest in productive assets, such as plant and equipment. They make these investment choices when they judge that their investments will increase the value of the firm by more than the cost of acquiring those funds from households. Firms also use such financial intermediaries as banks and insurance companies to raise capital, typically debt funding that ultimately comes from the savings of households, which are usually net accumulators of financial capital.

Microeconomics, although primarily focused on goods and factor markets, can contribute to the understanding of all types of markets (e.g., markets for financial securities).

EXAMPLE 1

Types of Markets

- 1 Which of the following markets is *least* accurately described as a factor market? The market for:
 - A land.
 - B assembly line workers.
 - C capital market securities.
- 2 Which of the following markets is *most* accurately defined as a goods market? The market for:
 - A companies.
 - B unskilled labor.
 - C legal and lobbying services.

Solution to 1:

C is correct.

Solution to 2:

C is correct.

3

BASIC PRINCIPLES AND CONCEPTS

In this reading, we will explore a model of household behavior that yields the consumer demand curve. **Demand**, in economics, is the willingness and ability of consumers to purchase a given amount of a good or service at a given price. **Supply** is the willingness of sellers to offer a given quantity of a good or service for a given price. Later, study on the theory of the firm will yield the supply curve.

The demand and supply model is useful in explaining how price and quantity traded are determined and how external influences affect the values of those variables. Buyers' behavior is captured in the demand function and its graphical equivalent, the demand curve. This curve shows both the highest price buyers are willing to pay

for each quantity, and the highest quantity buyers are willing and able to purchase at each price. Sellers' behavior is captured in the supply function and its graphical equivalent, the supply curve. This curve shows simultaneously the lowest price sellers are willing to accept for each quantity and the highest quantity sellers are willing to offer at each price.

If, at a given quantity, the highest price that buyers are willing to pay is equal to the lowest price that sellers are willing to accept, we say the market has reached its equilibrium quantity. Alternatively, when the quantity that buyers are willing and able to purchase at a given price is just equal to the quantity that sellers are willing to offer at that same price, we say the market has discovered the equilibrium price. So equilibrium price and quantity are achieved simultaneously, and as long as neither the supply curve nor the demand curve shifts, there is no tendency for either price or quantity to vary from their equilibrium values.

3.1 The Demand Function and the Demand Curve

We first analyze demand. The quantity consumers are willing to buy clearly depends on a number of different factors called variables. Perhaps the most important of those variables is the item's own price. In general, economists believe that as the price of a good rises, buyers will choose to buy less of it, and as its price falls, they buy more. This is such a ubiquitous observation that it has come to be called the **law of demand**, although we shall see that it need not hold in all circumstances.

Although a good's own price is important in determining consumers' willingness to purchase it, other variables also have influence on that decision, such as consumers' incomes, their tastes and preferences, the prices of other goods that serve as substitutes or complements, and so on. Economists attempt to capture all of these influences in a relationship called the **demand function**. (In general, a function is a relationship that assigns a unique value to a dependent variable for any given set of values of a group of independent variables.) We represent such a demand function in Equation 1:

$$Q_x^d = f(P_x, I, P_y, \dots) \quad (1)$$

where Q_x^d represents the quantity demanded of some good X (such as per household demand for gasoline in gallons per week), P_x is the price per unit of good X (such as \$ per gallon), I is consumers' income (as in \$1,000s per household annually), and P_y is the price of another good, Y . (There can be many other goods, not just one, and they can be complements or substitutes.) Equation 1 may be read, "Quantity demanded of good X depends on (is a function of) the price of good X , consumers' income, the price of good Y , and so on."

Often, economists use simple linear equations to approximate real-world demand and supply functions in relevant ranges. A hypothetical example of a specific demand function could be the following linear equation for a small town's per-household gasoline consumption per week, where P_y might be the average price of an automobile in \$1,000s:

$$Q_x^d = 8.4 - 0.4P_x + 0.06I - 0.01P_y \quad (2)$$

The signs of the coefficients on gasoline price (negative) and consumer's income (positive) are intuitive, reflecting, respectively, an inverse and a positive relationship between those variables and quantity of gasoline consumed. The negative sign on average automobile price may indicate that if automobiles go up in price, fewer will be purchased and driven; hence less gasoline will be consumed. As will be discussed later, such a relationship would indicate that gasoline and automobiles have a negative cross-price elasticity of demand and are thus complements.

To continue our example, suppose that the price of gasoline (P_x) is \$3 per gallon, per household income (I) is \$50,000, and the price of the average automobile (P_y) is \$20,000. Then this function would predict that the per-household weekly demand for gasoline would be 10 gallons: $8.4 - 0.4(3) + 0.06(50) - 0.01(20) = 8.4 - 1.2 + 3 - 0.2 = 10$, recalling that income and automobile prices are measured in thousands. Note that the sign on the own-price variable is negative, thus, as the price of gasoline rises, per household weekly consumption would decrease by 0.4 gallons for every dollar increase in gas price. **Own-price** is used by economists to underscore that the reference is to the price of a good itself and not the price of some other good.

In our example, there are three independent variables in the demand function, and one dependent variable. If any one of the independent variables changes, so does the value of quantity demanded. It is often desirable to concentrate on the relationship between the dependent variable and just one of the independent variables at a time, which allows us to represent the relationship between those two variables in a two-dimensional graph (at specific levels of the variables held constant). To accomplish this goal, we can simply hold the other two independent variables constant at their respective levels and rewrite the equation. In economic writing, this “holding constant” of the values of all variables except those being discussed is traditionally referred to by the Latin phrase *ceteris paribus* (literally, “all other things being equal” in the sense of “unchanged”). In this reading, we will use the phrase “holding all other things constant” as a readily understood equivalent for *ceteris paribus*.

Suppose, for example, that we want to concentrate on the relationship between the quantity demanded of the good and its own-price, P_x . Then we would hold constant the values of income and the price of good Y . In our example, those values are 50 and 20, respectively. So, by inserting the respective values, we would rewrite Equation 2 as

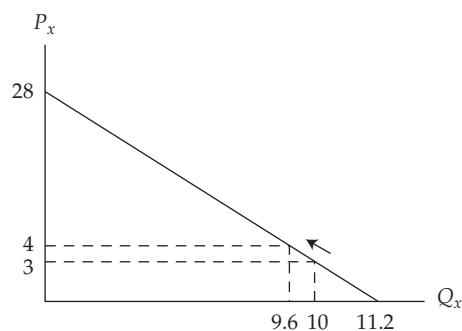
$$Q_x^d = 8.4 - 0.4P_x + 0.06(50) - 0.01(20) = 11.2 - 0.4P_x \quad (3)$$

Notice that income and the price of automobiles are not ignored; they are simply held constant, and they are “collected” in the new constant term, 11.2. Notice also that we can rearrange Equation 3, solving for P_x in terms of Q_x . This operation is called “inverting the demand function,” and gives us Equation 4. (You should be able to perform this algebraic exercise to verify the result.)

$$P_x = 28 - 2.5Q_x \quad (4)$$

Equation 4, which gives the per-gallon price of gasoline as a function of gasoline consumed per week, is referred to as the **inverse demand function**. We need to restrict Q_x in Equation 4 to be less than or equal to 11.2 so price is not negative. Henceforward we assume that the reader can work out similar needed qualifications to the valid application of equations. The graph of the inverse demand function is called the **demand curve**, and is shown in Exhibit 1.¹

¹ Following usual practice, here and in other exhibits we will show linear demand curves intersecting the quantity axis at a price of zero, which shows the intercept of the associated demand equation. Real-world demand functions may be non-linear in some or all parts of their domain. Thus, linear demand functions in practical cases are viewed as approximations to the true demand function that are useful for a relevant range of values. The relevant range would typically not include a price of zero, and the prediction for demand at a price of zero should not be viewed as usable.

Exhibit 1 Household Demand Curve for Gasoline

This demand curve is drawn with price on the vertical axis and quantity on the horizontal axis. Depending on how we interpret it, the demand curve shows either the highest quantity a household would buy at a given price or the highest price it would be willing to pay for a given quantity. In our example, at a price of \$3 per gallon households would each be willing to buy 10 gallons per week. Alternatively, the highest price they would be willing to pay for 10 gallons per week is \$3 per gallon. Both interpretations are valid, and we will be thinking in terms of both as we proceed. If the price were to rise by \$1, households would reduce the quantity they each bought by 0.4 units to 9.6 gallons. We say that the slope of the demand curve is $1/-0.4$, or -2.5 . Slope is always measured as “rise over run,” or the change in the vertical variable divided by the change in the horizontal variable. In this case, the slope of the demand curve is $\Delta P/\Delta Q$, where “ Δ ” stands for “the change in.” The change in price was \$1, and it is associated with a change in quantity of negative 0.4.

3.2 Changes in Demand vs. Movements along the Demand Curve

As we just saw, when own-price changes, quantity demanded changes. This change is called a movement along the demand curve or a change in quantity demanded, and it comes only from a change in own price.

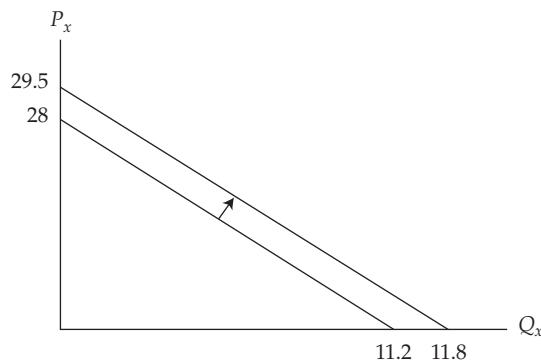
Recall that to draw the demand curve, though, we had to hold everything *except* quantity and own-price constant. What would happen if income were to change by some amount? Suppose that household income rose by \$10,000 per year to a value of 60. Then the value of Equation 3 would change to

$$Q_x^d = 8.4 - 0.4P_x + 0.06(60) - 0.01(20) = 11.8 - 0.4P_x \quad (5)$$

and Equation 4 would become the new inverse demand function:

$$P_x = 29.5 - 2.5Q_x \quad (6)$$

Notice that the slope has remained constant, but the intercepts have both increased, resulting in an outward shift in the demand curve, as shown in Exhibit 2.

Exhibit 2 Household Demand Curve for Gasoline before and after Change in Income


In general, the only thing that can cause a movement along the demand curve is a change in a good's own-price. A change in the value of any *other* variable will shift the entire demand curve. The former is referred to as a *change in quantity demanded*, and the latter is referred to as a *change in demand*.

More importantly, the shift in demand was both a *vertical* shift upward and a *horizontal* shift to the right. That is to say, for any given quantity, the household is now willing to pay a higher price; and at any given price, the household is now willing to buy a greater quantity. Both interpretations of the shift in demand are valid.

EXAMPLE 2
Representing Consumer Buying Behavior with a Demand Function and Demand Curve

An individual consumer's monthly demand for downloadable e-books is given by the equation

$$Q_{eb}^d = 2 - 0.4P_{eb} + 0.0005I + 0.15P_{hb}$$

where Q_{eb}^d equals the number of e-books demanded each month, P_{eb} equals the price of e-books, I equals the household monthly income, and P_{hb} equals the price of hardbound books, per unit. Notice that the sign on the price of hardbound books is positive, indicating that when hardbound books increase in price, more e-books are purchased; thus, according to this equation, the two types of books are substitutes. Assume that the price of e-books is €10.68, household income is €2,300, and the price of hardbound books is €21.40.

- 1 Determine the number of e-books demanded by this household each month.
- 2 Given the values for I and P_{hb} , determine the inverse demand function.
- 3 Determine the slope of the demand curve for e-books.
- 4 Calculate the vertical intercept (price-axis intercept) of the demand curve if income increases to €3000 per month.

Solution to 1:

Insert given values into the demand function and calculate quantity:

$$Q_{eb}^d = 2 - 0.4(10.68) + 0.0005(2,300) + 0.15(21.40) = 2.088$$

Hence, the household will demand e-books at the rate of 2.088 books per month. Note that this rate is a flow, so there is no contradiction in there being a non-integer quantity. In this case, the outcome means that the consumer buys 23 e-books per 11 months.

Solution to 2:

We want to find the price–quantity relationship holding all other things constant, so first, insert values for I and P_{hb} into the demand function and collect the constant terms:

$$Q_{eb}^d = 2 - 0.4P_{eb} + 0.0005(2,300) + 0.15(21.40) = 6.36 - 0.4P_{eb}$$

Now solve for P_{eb} in terms of Q_{eb} : $P_{eb} = 15.90 - 2.5Q_{eb}$

Solution to 3:

Note from the inverse demand function above that when Q_{eb} rises by one unit, P_{eb} falls by 2.5 euros. So the slope of the demand curve is -2.5 , which is the coefficient on Q_{eb} in the inverse demand function. Note it is *not* the coefficient on P_{eb} in the demand function, which is -0.4 . It is the inverse of that coefficient.

Solution to 4:

In the demand function, change the value of I to 3,000 from 2,300 and collect constant terms:

$$Q_{eb}^d = 2 - 0.4P_{eb} + 0.0005(3,000) + 0.15(21.40) = 6.71 - 0.4P_{eb}$$

Now solve for P_{eb} : $P_{eb} = 16.78 - 2.5Q_{eb}$. The vertical intercept is 16.78. (Note that this increase in income has shifted the demand curve outward and upward but has not affected its slope, which is still -2.5 .)

3.3 The Supply Function and the Supply Curve

The willingness and ability to sell a good or service is called supply. In general, producers are willing to sell their product for a price as long as that price is at least as high as the cost to produce an additional unit of the product. It follows that the willingness to supply, called the **supply function**, depends on the price at which the good can be sold as well as the cost of production for an additional unit of the good. The greater the difference between those two values, the greater is the willingness of producers to supply the good.

In another reading, we will explore the cost of production in greater detail. At this point, we need to understand only the basics of cost. At its simplest level, production of a good consists of transforming inputs, or factors of production (such as land, labor, capital, and materials) into finished goods and services. Economists refer to the “rules” that govern this transformation as the **technology of production**. Because producers have to purchase inputs in factor markets, the cost of production depends on both the technology and the price of those factors. Clearly, willingness to supply is dependent on not only the price of a producer’s output, but also additionally on the prices (i.e., costs) of the inputs necessary to produce it. For simplicity, we can assume that the only input in a production process is labor that must be purchased in the labor market. The price of an hour of labor is the wage rate, or W . Hence, we can say that (for any given level of technology) the willingness to supply a good depends on the price of that good and the wage rate. This concept is captured in the following equation, which represents an individual seller’s supply function:

$$Q_x^s = f(P_x, W, \dots) \quad (7)$$

where Q_x^s is the quantity supplied of some good X , such as gasoline, P_x is the price per unit of good X , and W is the wage rate of labor in, say, dollars per hour. It would be read, "The quantity supplied of good X depends on (is a function of) the price of X (its "own" price), the wage rate paid to labor, etc."

Just as with the demand function, we can consider a simple hypothetical example of a seller's supply function. As mentioned earlier, economists often will simplify their analysis by using linear functions, although that is not to say that all demand and supply functions are necessarily linear. One hypothetical example of an individual seller's supply function for gasoline is given in Equation 8:

$$Q_x^s = -175 + 250P_x - 5W \quad (8)$$

Notice that this supply function says that for every increase in price of \$1, this seller would be willing to supply an additional 250 units of the good. Additionally, for every \$1 increase in wage rate that it must pay its laborers, this seller would experience an increase in marginal cost and would be willing to supply five fewer units of the good.

We might be interested in the relationship between only two of these variables, price and quantity supplied. Just as we did in the case of the demand function, we use the assumption of *ceteris paribus* and hold everything except own-price and quantity constant. In our example, we accomplish this by setting W to some value, say, \$15. The result is Equation 9:

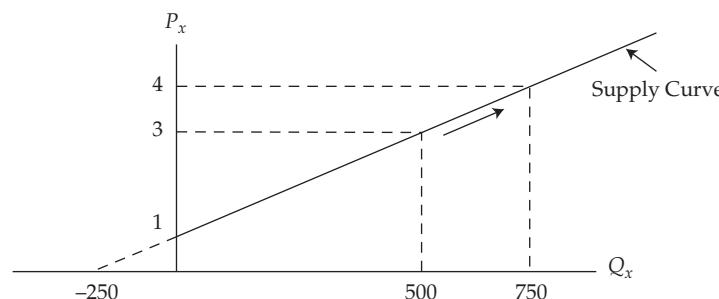
$$Q_x^s = -175 + 250P_x - 5(15) = -250 + 250P_x \quad (9)$$

in which only the two variables Q_x^s and P_x appear. Once again, we can solve this equation for P_x in terms of Q_x^s , which yields the *inverse supply function* in Equation 10:

$$P_x = 1 + 0.004Q_x \quad (10)$$

The graph of the inverse supply function is called the **supply curve**, and it shows simultaneously the highest quantity willingly supplied at each price and the lowest price willingly accepted for each quantity. For example, if the price of gasoline were \$3 per gallon, Equation 9 implies that this seller would be willing to sell 500 gallons per week. Alternatively, the lowest price she would accept and still be willing to sell 500 gallons per week would be \$3. Exhibit 3 represents our hypothetical example of an individual seller's supply curve of gasoline.

Exhibit 3 Individual Seller's Supply Curve for Gasoline



What does our supply function tell us will happen if the retail price of gasoline rises by \$1? We insert the new higher price of \$4 into Equation 8 and find that quantity supplied would rise to 750 gallons per week. The increase in price has enticed the seller to supply a greater quantity of gasoline per week than at the lower price.

3.4 Changes in Supply vs. Movements along the Supply Curve

As we saw earlier, a change in the (own) price of a product causes a change in the quantity of that good willingly supplied. A rise in price typically results in a greater quantity supplied, and a lower price results in a lower quantity supplied. Hence, the supply curve has a positive slope, in contrast to the negative slope of a demand curve. This positive relationship is often referred to as the **law of supply**.

What happens when a variable other than own-price takes on different values? We could answer this question in our example by assuming a different value for wage rate, say, \$20 instead of \$15. Recalling Equation 9, we would simply put in the higher wage rate and solve, yielding Equation 11.

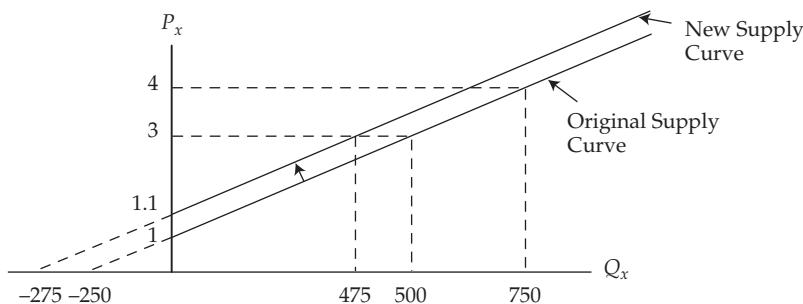
$$Q_x^s = -175 + 250P_x - 5(20) = -275 + 250P_x \quad (11)$$

This equation, too, can be solved for P_x , yielding the inverse supply function:

$$P_x = 1.1 + 0.004Q_x \quad (12)$$

Notice that the constant term has changed, but the slope has remained the same. The result is a shift in the entire supply curve, as illustrated in Exhibit 4:

Exhibit 4 Individual Seller's Supply Curve for Gasoline before and after Increase in Wage Rate



Notice that the supply curve has shifted both vertically upward and horizontally leftward as a result of the rise in the wage rate paid to labor. This change is referred to as a **change in supply**, as contrasted with a **change in quantity supplied** that would result only from a change in this product's own price. Now, at a price of 3, a lower quantity will be supplied: 475 instead of 500. Alternatively, in order to entice this seller to offer the same 500 gallons per week, the price would now have to be 3.1, up from 3 before the change. This increase in lowest acceptable price reflects the now higher marginal cost of production resulting from the increased input price the firm now must pay for labor.

To summarize, a change in the price of a good itself will result in a movement along the supply curve and a change in quantity supplied. A change in any variable other than own-price will cause a shift in the supply curve, called a change in supply. This distinction is identical to the case of demand curves.

EXAMPLE 3**Representing Seller Behavior with a Supply Function and Supply Curve**

An individual seller's monthly supply of downloadable e-books is given by the equation

$$Q_{eb}^s = -64.5 + 37.5P_{eb} - 7.5W$$

where Q_{eb}^s is number of e-books supplied each month, P_{eb} is price of e-books in euros, and W is the hourly wage rate in euros paid by e-book sellers to workers. Assume that the price of e-books is €10.68 and the hourly wage is €10.

- 1 Determine the number of e-books supplied each month.
- 2 Determine the inverse supply function for an individual seller.
- 3 Determine the slope of the supply curve for e-books.
- 4 Determine the new vertical intercept of the individual e-book supply curve if the hourly wage were to rise to €15 from €10.

Solution to 1:

Insert given values into the supply function and calculate the number of e-books:

$$Q_{eb}^s = -64.5 + 37.5(10.68) - 7.5(10) = 261$$

Hence, each seller would be willing to supply e-books at the rate of 261 per month.

Solution to 2:

Holding all other things constant, the wage rate is constant at €10, so we have

$$Q_{eb}^s = -64.5 + 37.5P_{eb} - 7.5(10) = -139.5 + 37.5P_{eb}$$

We now solve this for P_{eb} :

$$P_{eb} = 3.72 + 0.0267Q_{eb}$$

Solution to 3:

Note that when Q_{eb} rises by one unit, P_{eb} rises by 0.0267 euros, so the slope of the supply curve is 0.0267, which is the coefficient on Q_{eb} in the inverse supply function. Note that it is *not* 37.5.

Solution to 4:

In the supply function, increase the value of W to €15 from €10:

$$Q_{eb}^s = -64.5 + 37.5P_{eb} - 7.5(15) = -177 + 37.5P_{eb}$$

and invert by solving for P_{eb} :

$$P_{eb} = 4.72 + 0.0267Q_{eb}$$

The vertical intercept is now 4.72. Thus, an increase in the wage rate shifts the supply curve upward and to the left. This change is known as a decrease in supply because at each price the seller would be willing now to supply fewer e-books than before the increase in labor cost.

3.5 Aggregating the Demand and Supply Functions

We have explored the basic concept of demand and supply at the individual household and the individual supplier level. However, markets consist of collections of demanders and suppliers, so we need to understand the process of combining these individual agents' behavior to arrive at market demand and supply functions.

The process could not be more straightforward: simply add all the buyers together and add all the sellers together. Suppose there are 1,000 identical gasoline buyers in our hypothetical example, and they represent the total market. At, say, a price of \$3 per gallon, we find that one household would be willing to purchase 10 gallons per week (when income and price of automobiles are held constant at \$50,000 and \$20,000, respectively). So, 1,000 identical buyers would be willing to purchase 10,000 gallons collectively. It follows that to aggregate 1,000 buyers' demand functions, simply multiply each buyer's quantity demanded by 1,000:

$$Q_x^d = 1,000(8.4 - 0.4P_x + 0.06I - 0.01P_y) = 8,400 - 400P_x + 60I - 10P_y \quad (13)$$

where Q_x^d represents the market quantity demanded. Note that if we hold I and P_y at their same respective values of 50 and 20 as before, we can "collapse" the constant terms and write the following Equation 14:

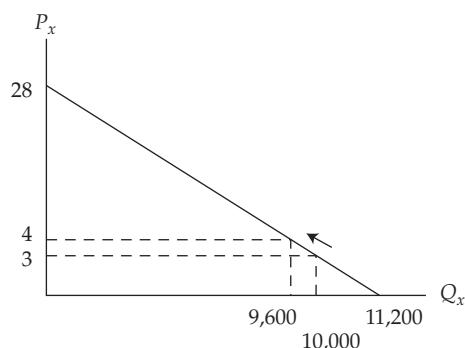
$$Q_x^d = 11,200 - 400P_x \quad (14)$$

Equation 14 is just Equation 3 (an individual household's demand function) multiplied by 1,000 households (Q_x^d represents thousands of gallons per week). Again, we can solve for P_x to obtain the market inverse demand function:

$$P_x = 28 - 0.0025Q_x \quad (15)$$

The market demand curve is simply the graph of the market inverse demand function, as shown in Exhibit 5.

Exhibit 5 Aggregate Weekly Market Demand for Gasoline as the Quantity Summation of all Households' Demand Curves



It is important to note that the aggregation process sums all individual buyers' *quantities*, not the *prices* they are willing to pay—that is, we multiplied the demand function, *not* the inverse demand function, by the number of households. Accordingly, the market demand curve has the exact same price intercept as each individual household's demand curve. If, at a price of \$28, a single household would choose to buy zero, then it follows that 1,000 identical households would choose, in aggregate, to buy zero as well. On the other hand, if each household chooses to buy 10 at a price of \$3,

then 1,000 identical households would choose to buy 10,000, as shown in Exhibit 5. Hence, we say that all individual demand curves *horizontally* (quantities), not *vertically* (prices), are added to arrive at the market demand curve.

Now that we understand the aggregation of demanders, the aggregation of suppliers is simple: We do exactly the same thing. Suppose, for example, that there are 20 identical sellers with the supply function given by Equation 8. To arrive at the market supply function, we simply multiply by 20 to obtain:

$$Q_x^s = 20(-175 + 250P_x - 5W) = -3,500 + 5,000P_x - 100W \quad (16)$$

And, if we once again assume W equals \$15, we can “collapse” the constant terms, yielding

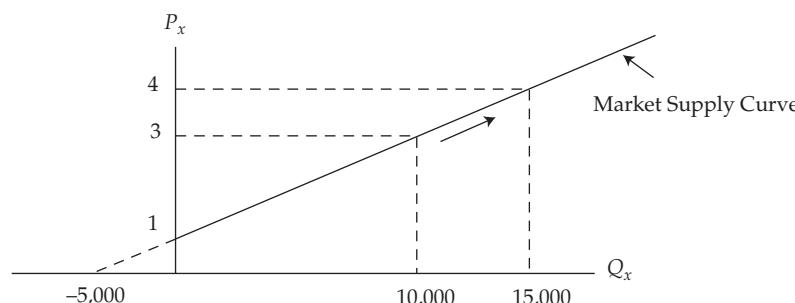
$$Q_x^s = 20[-175 + 250P_x - 5(15)] = -5,000 + 5,000P_x \quad (17)$$

which can be inverted to yield the market inverse supply function:

$$P_x = 1 + 0.0002Q_x \quad (18)$$

Graphing the market inverse supply function yields the market supply curve in Exhibit 6:

Exhibit 6 Aggregate Market Supply as the Quantity Summation of Individual Sellers' Supply Curves



We saw from the individual seller's supply curve in Exhibit 3 that at a price of \$3, an individual seller would willingly offer 500 gallons of gasoline. It follows, as shown in Exhibit 6, that a group of 20 sellers would offer 10,000 gallons per week. Accordingly, at each price, the market quantity supplied is just 20 times as great as the quantity supplied by each seller. We see, as in the case of demand curves, that the market supply curve is simply the horizontal summation of all individual sellers' supply curves.

EXAMPLE 4

Aggregating Demand Functions

An individual consumer's monthly demand for downloadable e-books is given by the equation

$$Q_{eb}^d = 2 - 0.4P_{eb} + 0.0005I + 0.15P_{hb}$$

where Q_{eb}^d equals the number of e-books demanded each month, P_{eb} is the price of e-books in euros, I equals the household monthly income, and P_{hb} equals the price of hardbound books, per unit. Assume that household income is €2,300, and the price of hardbound books is €21.40. The market consists of 1,000 identical consumers with this demand function.

- 1 Determine the market aggregate demand function.
- 2 Determine the inverse market demand function.
- 3 Determine the slope of the market demand curve.

Solution to 1:

Aggregating over the total number of consumers means summing up their demand functions (in the quantity direction). In this case, there are 1,000 consumers with identical individual demand functions, so multiply the entire function by 1,000:

$$\begin{aligned} Q_{eb} &= 1,000(2 - 0.4P_{eb} + 0.0005I + 0.15P_{hb}) \\ &= 2,000 - 400P_{eb} + 0.5I + 150P_{hb} \end{aligned}$$

Solution to 2:

Holding I constant at a value of €2,300 and P_{hb} constant at a value of €21.40, we find

$$Q_{eb} = 2,000 - 400P_{eb} + 0.5(2300) + 150(21.40) = 6,360 - 400P_{eb}$$

Now solve for $P_{eb} = 15.90 - 0.0025Q_{eb}$

Solution to 3:

The slope of the market demand curve is the coefficient on Q_{eb} in the inverse demand function, which is -0.0025 .

EXAMPLE 5

Aggregating Supply Functions

An individual seller's monthly supply of downloadable e-books is given by the equation

$$Q_{eb}^s = -64.5 + 37.5P_{eb} - 7.5W$$

where Q_{eb}^s is number of e-books supplied, P_{eb} is the price of e-books in euros, and W is the wage rate in euros paid by e-book sellers to laborers. Assume that the price of e-books is €10.68 and wage is €10. The supply side of the market consists of a total of eight identical sellers in this competitive market.

- 1 Determine the market aggregate supply function.
- 2 Determine the inverse market supply function.
- 3 Determine the slope of the aggregate market supply curve.

Solution to 1:

Aggregating supply functions means summing up the quantity supplied by all sellers. In this case, there are eight identical sellers, so multiply the individual seller's supply function by eight:

$$Q_{eb}^s = 8(-64.5 + 37.5P_{eb} - 7.5W) = -516 + 300P_{eb} - 60W$$

Solution to 2:

Holding W constant at a value of €10, insert that value into the aggregate supply function and then solve for P_{eb} to find the inverse supply function:

$$Q_{eb} = -1,116 + 300P_{eb}$$

$$\text{Inverting, } P_{eb} = 3.72 + 0.0033Q_{eb}$$

Solution to 3:

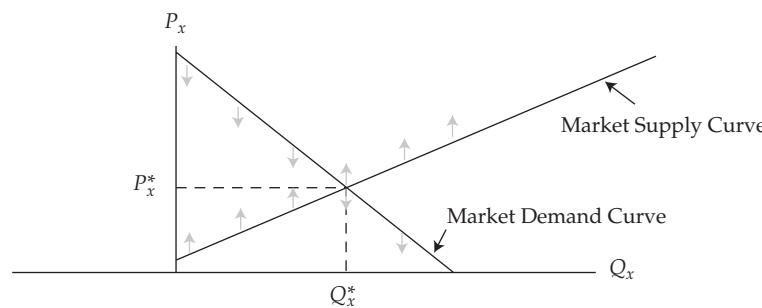
The slope of the supply curve is the coefficient on Q_{eb} in the inverse supply function, which is 0.0033.

3.6 Market Equilibrium

An important concept in the market model is **market equilibrium**, defined as the condition in which the quantity willingly offered for sale by sellers at a given price is just equal to the quantity willingly demanded by buyers at that same price. When that condition is met, we say that the market has discovered its equilibrium price. An alternative and equivalent condition of equilibrium occurs at that quantity at which the highest price a buyer is willing to pay is just equal to the lowest price a seller is willing to accept for that same quantity.

As we have discovered in the earlier sections, the demand curve shows (for given values of income, other prices, etc.) an infinite number of combinations of prices and quantities that satisfy the demand function. Similarly, the supply curve shows (for given values of input prices, etc.) an infinite number of combinations of prices and quantities that satisfy the supply function. Equilibrium occurs at the unique combination of price and quantity that simultaneously satisfies *both* the market demand function and the market supply function. Graphically, it is the intersection of the demand and supply curves as shown in Exhibit 7.

Exhibit 7 Market Equilibrium Price and Quantity as the Intersection of Demand and Supply



In Exhibit 7, the shaded arrows indicate, respectively, that buyers will be willing to pay any price at or below the demand curve (indicated by ↓), and sellers are willing to accept any price at or above the supply curve (indicated by ↑). Notice that for

quantities less than Q_x^* , the highest price buyers are willing to pay exceeds the lowest price sellers are willing to accept, as indicated by the shaded arrows. But for all quantities above Q_x^* , the lowest price willingly accepted by sellers is greater than the highest price willingly offered by buyers. Clearly, trades will not be made beyond Q_x^* .

Algebraically, we can find equilibrium price by setting the demand function equal to the supply function and solving for price. Recall that in our hypothetical example of a local gasoline market, the demand function was given by $Q_x^d = f(P_x, I, P_y)$, and the supply function was given by $Q_x^s = f(P_x, W)$. Those expressions are called **behavioral equations** because they model the behavior of, respectively, buyers and sellers. Variables *other* than own price and quantity are determined outside of the demand and supply model of this particular market. Because of that, they are called **exogenous variables**. Price and quantity, however, are determined within the model for this particular market and are called **endogenous variables**. In our simple example, there are three exogenous variables (I , P_y , and W) and three endogenous variables: P_x , Q_x^d , and Q_x^s . Hence, we have a system of two equations and three unknowns. We need another equation to solve this system. That equation is called the **equilibrium condition**, and it is simply $Q_x^d = Q_x^s$.

Continuing with our hypothetical examples, we could assume that income equals \$50 (thousand, per year), the price of automobiles equals \$20 (thousand, per automobile), and the hourly wage equals \$15. In this case, our equilibrium condition can be represented by setting Equation 14 equal to Equation 17:

$$11,200 - 400P_x = -5,000 + 5,000P_x \quad (19)$$

and solving for equilibrium, $P_x = 3$.

Equivalently, we could have equated the inverse demand function to the inverse supply function (Equations 15 and 18, respectively)

$$28 - 0.0025Q_x = 1 + 0.0002Q_x \quad (20)$$

and solved for equilibrium, $Q_x = 10,000$. That is to say, for the given values of I and W , the unique combination of price and quantity of gasoline that results in equilibrium is (3, 10,000).

Note that our system of equations requires explicit values for the exogenous variables to find a unique equilibrium combination of price and quantity. Conceptually, the values of the exogenous variables are being determined in other markets, such as the markets for labor, automobiles, and so on, whereas the price and quantity of gasoline are being determined in the gasoline market. When we concentrate on one market, taking values of exogenous variables as given, we are engaging in what is called **partial equilibrium analysis**. In many cases, we can gain sufficient insight into a market of interest without addressing feedback effects to and from all the other markets that are tangentially involved with this one. At other times, however, we need explicitly to take account of all the feedback mechanisms that are going on in all markets simultaneously. When we do that, we are engaging in what is called **general equilibrium analysis**. For example, in our hypothetical model of the local gasoline market, we recognize that the price of automobiles, a complementary product, has an impact on the demand for gasoline. If the price of automobiles were to rise, people would tend to buy fewer automobiles and probably buy less gasoline. Additionally, though, the price of gasoline probably has an impact on the demand for automobiles that, in turn, can feed back to the gasoline market. Because we are positing a very local gasoline market, it is probably safe to ignore all the feedback effects, but if we are modeling the national markets for gasoline and automobiles, a general equilibrium model might be warranted.

EXAMPLE 6**Finding Equilibrium by Equating Demand and Supply**

In the local market for e-books, the aggregate demand is given by the equation

$$Q_{eb}^d = 2,000 - 400P_{eb} + 0.5I + 150P_{hb}$$

and the aggregate supply is given by the equation

$$Q_{eb}^s = -516 + 300P_{eb} - 60W$$

where Q_{eb} is quantity of e-books, P_{eb} is the price of an e-book, I is household income, W is wage rate paid to e-book laborers, and P_{hb} is the price of a hard-bound book. Assume I is €2,300, W is €10, and P_{hb} is €21.40. Determine the equilibrium price and quantity of e-books in this local market.

Solution:

Market equilibrium occurs when quantity demanded is equal to quantity supplied, so set $Q_{eb}^d = Q_{eb}^s$ after inserting the given values for the exogenous variables:

$$\begin{aligned} 2,000 - 400P_{eb} + 0.5(2,300) + 150(21.4) &= -516 + 300P_{eb} - 60(10) \\ 6,360 - 400P_{eb} &= -1,116 + 300P_{eb}, \end{aligned}$$

which implies that $P_{eb} = €10.68$, and $Q_{eb} = 2,088$.

3.7 The Market Mechanism: Iterating toward Equilibrium—or Not

It is one thing to define equilibrium as we have done, but we should also understand the mechanism for reaching equilibrium. That mechanism is what takes place when the market is *not* in equilibrium. Consider our hypothetical example. We found that the equilibrium price was 3, but what would happen if, by some chance, price was actually equal to 4? To find out, we need to see how much buyers would demand at that price and how much sellers would offer to sell by inserting 4 into the demand function and into the supply function.

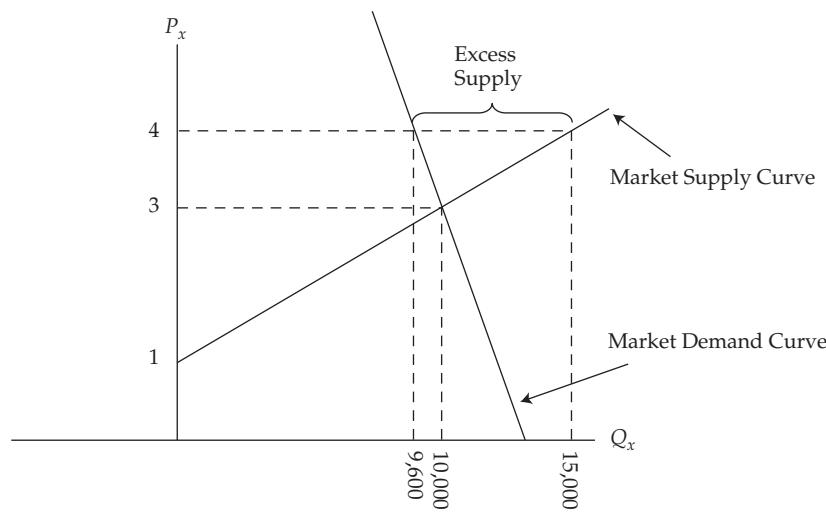
In the case of quantity demanded, we find that

$$Q_x^d = 11,200 - 400(4) = 9,600 \quad (21)$$

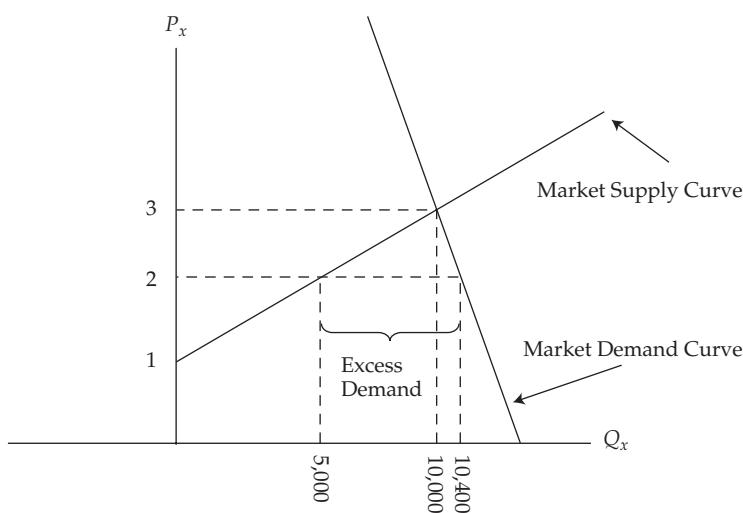
and in the case of quantity supplied,

$$Q_x^s = -5,000 + 5,000(4) = 15,000 \quad (22)$$

Clearly, the quantity supplied is greater than the quantity demanded, resulting in a condition called **excess supply**, as illustrated in Exhibit 8. In our example, there are 5,400 more units of this good offered for sale at a price of 4 than are demanded at that price.

Exhibit 8 Excess Supply as a Consequence of Price above Equilibrium Price

Alternatively, if the market was presented with a price that was too low, say 2, then by inserting the price of 2 into Equations 21 and 22, we find that buyers are willing to purchase 5,400 *more* units than sellers are willing to offer. This result is shown in Exhibit 9.

Exhibit 9 Excess Demand as a Consequence of Price below Equilibrium Price

To reach equilibrium, price must adjust until there is neither an excess supply nor an excess demand. That adjustment is called the **market mechanism**, and it is characterized in the following way: In the case of excess supply, price will fall; in the case of excess demand, price will rise; and in the case of neither excess supply nor excess demand, price will not change.

EXAMPLE 7**Identifying Excess Demand or Excess Supply at a Non-equilibrium Price**

In the local market for e-books, the aggregate demand is given by the equation

$$Q_{eb}^d = 6,360 - 400P_{eb}$$

and the aggregate supply by the equation

$$Q_{eb}^s = -1,116 + 300P_{eb}$$

- 1 Determine the amount of excess demand or supply if price is €12.
- 2 Determine the amount of excess demand or supply if price is €8.

Solution to 1:

Insert the presumed price of €12 into the demand function to find $Q_{eb}^d = 6,360 - 400(12) = 1,560$. Insert a price of €12 into the supply function to find $Q_{eb}^s = -1,116 + 300(12) = 2,484$. Because quantity supplied is greater than quantity demanded at the €12 price, there is an excess supply equal to $2,484 - 1,560 = 924$.

Solution to 2:

Insert the presumed price of €8 into the demand function to find $Q_{eb}^d = 6,360 - 400(8) = 3,160$. Insert a price of €8 into the supply function to find $Q_{eb}^s = -1,116 + 300(8) = 1,284$. Because quantity demanded is greater than quantity supplied at the €8 price, there is an excess demand equal to $3,160 - 1,284 = 1,876$.

It might be helpful to consider the following process in our hypothetical market. Suppose that some neutral agent or referee were to display a price for everyone in the market to observe. Then, given that posted price, we would ask each potential buyer to write down on a slip of paper a quantity that he/she would be willing and able to purchase at that price. At the same time, each potential seller would write down a quantity that he/she would be willing to sell at that price. Those pieces of paper would be submitted to the referee who would then calculate the total quantity demanded and the total quantity supplied at that price. If the two sums are identical, the slips of paper would essentially become contracts that would be executed, and the session would be concluded by buyers and sellers actually trading at that price. If there was an excess supply, however, the referee's job would be to discard the earlier slips of paper and display a price lower than before. Alternatively, if there was an excess demand at the original posted price, the referee would discard the slips of paper and post a higher price. This process would continue until the market reached an equilibrium price at which the quantity willingly offered for sale would just equal the quantity willingly purchased. In this way, the market could tend to move toward equilibrium.²

It is not really necessary for a market to have such a referee for it to operate *as if* it had one. Experimental economists have simulated markets in which subjects (usually college students) are given an “order” either to purchase or sell some amount of a

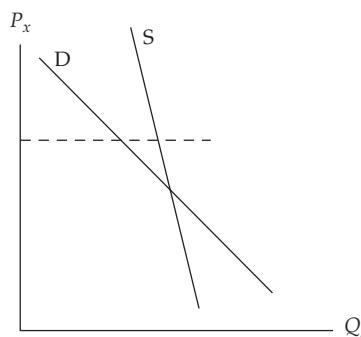
² The process described is known among economists as Walrasian *tâtonnement*, after the French economist Léon Walras (1834–1910). “*Tâtonnement*” means roughly, “searching,” referring to the mechanism for establishing the equilibrium price.

commodity for a price either no higher (in the case of buyers) or no lower (in the case of sellers) than a set dollar limit. Those limits are distributed among market participants and represent a positively sloped supply curve and a negatively sloped demand curve. The goal for buyers is to buy at a price as far below their limit as possible, and for sellers to sell at a price as far above their limit as possible. The subjects are then allowed to interact in a simulated trading pit by calling out willingness to buy or sell. When two participants come to an agreement on a price, that trade is then reported to a recorder who displays the terms of the deal. Traders are then allowed to observe current prices as they continue to search for a buyer or seller. It has consistently been shown in experiments that this mechanism of open outcry buying and selling (historically, one of the oldest mechanisms used in trading securities) soon converges to the theoretical equilibrium price and quantity inherent in the underlying demand and supply curves used to set the respective sellers' and buyers' limit prices.

In our hypothetical example of the gasoline market, the supply curve is positively sloped, and the demand curve is negatively sloped. In that case, the market mechanism would tend to reach an equilibrium whenever price was accidentally "bumped" away from it. We refer to such an equilibrium as being **stable** because whenever price is disturbed away from equilibrium, it tends to converge back to that equilibrium.³ It is possible, however, for this market mechanism to result in an unstable equilibrium. Suppose that not only the demand curve has a negative slope but also the supply curve has a negatively sloped segment. For example, at some level of wages, a wage increase might cause workers to supply fewer hours of work if satisfaction ("utility") gained from an extra hour of leisure is greater than the satisfaction obtained from an extra hour of work. Then two possibilities could result, as shown in Panels A and B of Exhibit 10.

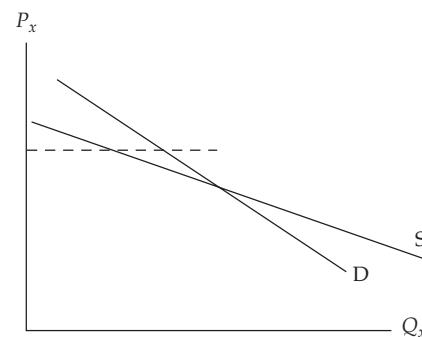
Exhibit 10 Stability of Equilibria: I

Panel A



Note: If supply intersects demand from above, equilibrium is dynamically stable.

Panel B



Note: If supply intersects demand from below, equilibrium is dynamically unstable.

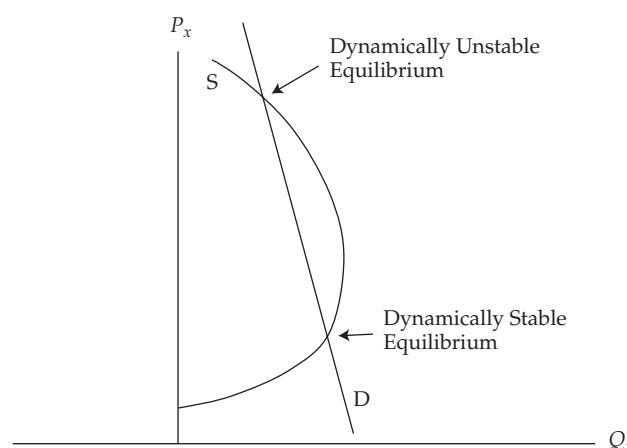
Notice that in Panel A both demand (D) and supply (S) are negatively sloped, but S is steeper and intersects D from above. In this case, if price is above equilibrium, there will be excess supply and the market mechanism will adjust price downward toward equilibrium. In Panel B, D is steeper, which results in S intersecting D from below. In this case, at a price above equilibrium there will be excess demand, and the market mechanism will dictate that price should *rise*, thus leading away from equilibrium.

³ In the same sense, equilibrium may sometimes also be referred to as being *dynamically stable*. Similarly, *unstable* or *dynamically unstable* may be used in the sense introduced later.

This equilibrium would be considered **unstable**. If price were accidentally displayed above the equilibrium price, the mechanism would not cause price to converge to that equilibrium, but instead to soar above it because there would be excess demand at that price. In contrast, if price were accidentally displayed below equilibrium, the mechanism would force price even further below equilibrium because there would be excess supply.

If supply were non-linear, there could be multiple equilibria, as shown in Exhibit 11.

Exhibit 11 Stability of Equilibria: II



Note: Multiple equilibria (stable and unstable) can result from nonlinear supply curves.

Note that there are two combinations of price and quantity that would equate quantity supplied and demanded, hence two equilibria. The lower-priced equilibrium is stable, with a positively sloped supply curve and a negatively sloped demand curve. However, the higher-priced equilibrium is unstable because at a price above that equilibrium price there would be excess demand, thus driving price even higher. At a price below that equilibrium there would be excess supply, thus driving price even lower toward the lower-priced equilibrium, which is a stable equilibrium.

Observation suggests that most markets are characterized by stable equilibria. Prices do not often shoot off to infinity or plunge toward zero. However, occasionally we do observe price bubbles occurring in real estate, securities, and other markets. It appears that prices can behave in ways that are not ultimately sustainable in the long run. They may shoot up for a time but ultimately, if they do not reflect actual valuations, the bubble can burst resulting in a “correction” to a new equilibrium.

As a simple approach to understanding bubbles, consider a case in which buyers and sellers base their expectations of future prices on the rate of change of current prices: if price rises, they take that as a sign that price will rise even further. Under these circumstances, if buyers see an increase in price today, they might actually shift the demand curve to the right, desiring to buy more at each price today because they expect to have to pay more in the future. Alternately, if sellers see an increase in today's price as evidence that price will be even higher in the future, they are reluctant to sell today as they hold out for higher prices tomorrow, and that would shift the supply curve to the left. With a rightward shift in demand and a leftward shift in supply, buyers' and sellers' expectations about price are confirmed and the process begins again. This scenario could result in a bubble that would inflate until someone decides that such high prices can no longer be sustained. The bubble bursts and price plunges.

3.8 Auctions as a Way to Find Equilibrium Price

Sometimes markets really do use auctions to arrive at equilibrium price. Auctions can be categorized into two types depending on whether the value of the item being sold is the same for each bidder or is unique to each bidder. The first case is called a **common value auction** in which there is some actual common value that will ultimately be revealed after the auction is settled. Prior to the auction's settlement, however, bidders must estimate that true value. An example of a common value auction would be bidding on a jar containing many coins. Each bidder could estimate the value; but until someone buys the jar and actually counts the coins, no one knows with certainty the true value. In the second case, called a **private value auction**, each buyer places a subjective value on the item, and in general their values differ. An example might be an auction for a unique piece of art that buyers are hoping to purchase for their own personal enjoyment, not primarily as an investment to be sold later.

Auctions also differ according to the mechanism used to arrive at a price and to determine the ultimate buyer. These mechanisms include the ascending price (or English) auction, the first price sealed bid auction, the second price sealed bid (or Vickery) auction, and the descending price (or Dutch) auction.

Perhaps the most familiar auction mechanism is the **ascending price auction** in which an auctioneer is selling a single item in a face-to-face arena where potential buyers openly reveal their willingness to buy the good at prices that are called out by an auctioneer. The auctioneer begins at a low price and easily elicits nods from buyers. He then raises the price incrementally. In a common value auction, buyers can sometimes learn something about the true value of the item being auctioned from observing other bidders. Ultimately bidders with different maximum amounts they are willing to pay for the item, called **reservation prices**, begin to drop out of the bidding as price rises above their respective reservation prices.⁴ Finally, only one bidder is left (who has outbid the bidder with the second highest valuation) and the item is sold to that bidder for his bid price.

Sometimes sellers offer a common value item, such as an oil or timber lease, in a **sealed bid auction**. In this case, bids are elicited from potential buyers, but there is no ability to observe bids by other buyers until the auction has ended. In the **first price sealed bid auction**, the envelopes containing bids are opened simultaneously and the item is sold to the highest bidder for the actual bid price. Consider an oil lease being auctioned by the government. The highest bidder will pay his bid price but does not know with certainty the profitability of the asset on which he is bidding. The profits that are ultimately realized will be learned only after a successful bidder buys and exploits the asset. Bidders each have some expected value they place on the oil lease, and those values can vary among bidders. Typically, some overly optimistic bidders will value the asset higher than its ultimate realizable value, and they might submit bids above that true value. Because the highest bidder wins the auction and must pay his full bid price, he may find that he has fallen prey to the **winner's curse** of having bid more than the ultimate value of the asset. The "winner" in this case will lose money because he has paid more than the value of the asset being auctioned. In recognition of the possibility of being overly optimistic, bidders might bid very conservatively below their expectation of the true value. If all bidders react in this way, the seller might end up with a low sale price.

If the item being auctioned is a private value item, then there is no danger of the winner's curse (no one would bid more than their own true valuation). But bidders try to guess the reservation prices of other bidders, so the most successful winning bidder would bid a price just above the reservation price of the second-highest bidder.

⁴ The term *reservation price* is also used to refer to the minimum price the seller of the auctioned item is willing to accept.

This bid will be below the true reservation price of the highest bidder, resulting in a “bargain” for the highest bidder. To induce each bidder to reveal their true reservation price, sellers can use the **second price sealed bid** mechanism (also known as a Vickery auction). In this mechanism, the bids are submitted in sealed envelopes and opened simultaneously. The winning buyer is the one who submitted the highest bid, but the price she pays is not equal to her own bid. She pays a price equal to the second-highest bid. The optimal strategy for any bidder in such an auction is to bid her actual reservation price, so the second price sealed bid auction induces buyers to reveal their true valuation of the item. It is also true that if the bidding increments are small, the second price sealed bid auction will yield the same ultimate price as the ascending price auction.

Yet another type of auction is called a **descending price auction** or **Dutch auction** in which the auctioneer begins at a very high price—a price so high that no bidder is believed to be willing to pay it.⁵ The auctioneer then lowers the called price in increments until there is a willing buyer of the item being sold. If there are many bidders, each with a different reservation price and a unit demand, then each has a perfectly vertical demand curve at one unit and a height equal to his reservation price. For example, suppose the highest reservation price is equal to \$100. That person would be willing to buy one unit of the good at a price no higher than \$100. Suppose each subsequent bidder also has a unit demand and a reservation price that falls, respectively, in increments of \$1. The market demand curve would be a negatively sloped step function; that is, it would look like a stair step, with the width of each step being one unit and the height of each step being \$1 lower than the preceding step. For example, at a price equal to \$90, 11 people would be willing to buy one unit of the good. If the price were to fall to \$89, then the quantity demanded would be 12, and so on.

In the Dutch auction, the auctioneer would begin with a price above \$100 and then lower it by increments until the highest reservation price bidder would purchase the unit. Again, the supply curve for this single unit auction would be vertical at one unit, although there might be a seller reserve price that would form the lower bound on the supply curve at that reserve price.

A traditional Dutch auction as just described could be conducted in a single unit or multiple unit format. With a multiple unit format, the price quoted by the auctioneer would be the per-unit price and a winning bidder could take fewer units than all the units for sale. If the winning bidder took fewer than all units for sale, the auctioneer would then lower the price until all units for sale were sold; thus transactions could occur at multiple prices. Modified Dutch auctions (frequently also called simply “Dutch Auctions” in practice) are commonly used in securities markets; the modifications often involve establishing a single price for all purchasers. As implemented in share repurchases, the company stipulates a range of acceptable prices at which the company would be willing to repurchase shares from existing shareholders. The auction process is structured to uncover the minimum price at which the company can buy back the desired number of shares, with the company paying that price to all qualifying bids. For example, if the share price is €25 per share, the company might offer to repurchase 3 million shares in a range of €26 to €28 per share. Each shareholder would then indicate the number of shares and the lowest price at which he or she would be willing to sell. The company would then begin to qualify bids beginning with those shareholders who submitted bids at €26 and continue to qualify bids at higher prices until 3 million shares had been qualified. In our example, that price might be €27. Shareholders who bid between €26 and €27, inclusive, would then be paid €27 per share for their shares.

⁵ The historical use of this auction type for flower auctions in the Netherlands explains the name.

Another Dutch auction variation, also involving a single price and called a **single price auction**, is used in selling US Treasury securities.⁶ The single price Treasury bill auction operates as follows: The Treasury announces that it will auction 26-week T-bills with an offering amount of, say, \$90 billion with both competitive and non-competitive bidding. Non-competitive bidders state the total face value they are willing to purchase at the ultimate price (yield) that clears the market (i.e., sells all of the securities offered), whatever that turns out to be. Competitive bidders each submit a total face value amount and the price at which they are willing to purchase those bills. The Treasury then ranks those bids in ascending order of yield (i.e., descending order of price) and finds the yield at which the total \$90 billion offering amount would be sold. If the offering amount is just equal to the total face value bidders are willing to purchase at that yield, then all the T-bills are sold for that single yield. If there is excess demand at that yield, then bidders would each receive a proportionately smaller total than they offered.

As an example, suppose the following table shows the prices and the offers from competitive bidders for a variety of prices, as well as the total offers from non-competitive bidders, assumed to be \$15 billion:

Discount Rate Bid (%)	Bid Price per \$100	Competitive Bids (\$ billions)	Cumulative Competitive Bids (\$ billions)	Non-competitive Bids (\$ billions)	Total Cumulative Bids (\$ billions)
0.1731	99.91250	10	10	15	25
0.1741	99.91200	15	25	15	40
0.1751	99.91150	20	45	15	60
0.1760	99.91100	12	57	15	72
0.1770	99.91050	10	67	15	82
0.1780	99.91000	5	72	15	87
0.1790	99.90950	10	82	15	97

At yields below 0.1790 percent (prices above 99.90950), there is still excess supply. But at that yield, more bills are demanded than the \$90 billion face value of the total offer amount. The clearing yield would be 0.1790 percent (a price of 99.9095 per \$100 of face value), and all sales would be made at that single yield. All the non-competitive bidders would have their orders filled at the clearing price, as well as all bidders who bid above that price. The competitive bidders who offered a price of 99.9095 would have 30 percent of their order filled at that price because it would take only 30 percent of the \$10 billion (\$90 billion – \$87 billion offered = \$3 billion, or 30 percent of \$10 billion) demanded at that price to complete the \$90 billion offer amount. That is, by filling 30 percent of the competitive bids at a price of 99.9095, the cumulative competitive bids would sum to \$75 billion. This amount plus the \$15 billion non-competitive bids adds up to \$90 billion.

EXAMPLE 8

Auctioning Treasury Bills with a Single Price Auction

The US Treasury offers to sell \$115 billion of 52-week T-bills and requests competitive and non-competitive bids. Non-competitive bids total \$10 billion, and competitive bidders in descending order of offer price are as given in the table below:

⁶ Historically, the US Treasury has also used multiple price auctions and in the euro area multiple price auctions are widely used. See http://www.dsti.nl/english/Subjects/Auction_methods for more information.

Discount Rate Bid (%)	Bid Price per \$100	Competitive Bids (\$ billions)	Cumulative Competitive Bids (\$ billions)	Non-competitive Bids (\$ billions)	Total Cumulative Bids (\$ billions)
0.1575	99.8425	12			
0.1580	99.8420	20			
0.1585	99.8415	36			
0.1590	99.8410	29			
0.1595	99.8405	5			
0.1600	99.8400	15			
0.1605	99.8395	10			

- 1 Determine the winning price if a single price Dutch auction is used to sell these T-bills.
- 2 For those bidders at the winning price, what percentage of their order would be filled?

Solution to 1:

Enter the non-competitive quantity of \$10 billion into the table. Then find the cumulative competitive bids and the total cumulative bids in the respective columns:

Bid Price per \$100	Competitive Bids (\$ billions)	Cumulative Competitive Bids (\$ billions)	Non-competitive Bids (\$ billions)	Total Cumulative Bids (\$ billions)
99.8425	12	12	10	22
99.8420	20	32	10	42
99.8415	36	68	10	78
99.8410	29	97	10	107
99.8405	5	102	10	112
99.8400	15	117	10	127
99.8395	10	127	10	137

Note that at a bid price of 99.8400 there would be excess demand of \$12 billion (i.e., the difference between \$127 billion bid and \$115 billion offered), but at the higher price of 99.8405 there would be excess supply. So the winning bid would be at a price of 99.8400.

Solution to 2:

At a price of 99.8400, there would be \$15 billion more demanded than at 99.8405 (\$127 billion minus \$112 billion), and at 99.8405 there would be excess supply equal to \$3 billion. So the bidders at the winning bid would have only $\frac{1}{15}$, or 20 percent, of their orders filled.

3.9 Consumer Surplus—Value minus Expenditure

To this point, we have discussed the fundamentals of demand and supply curves and explained a simple model of how a market can be expected to arrive at an equilibrium combination of price and quantity. While it is certainly necessary for the analyst to understand the basic working of the market model, it is also crucial to have a sense of

why we might care whether the market tends toward equilibrium. This question moves us into the normative, or evaluative, consideration of whether market equilibrium is desirable in any social sense. In other words, is there some reasonable measure we can apply to the outcome of a competitive market that enables us to say whether that outcome is socially desirable? Economists have developed two related concepts called **consumer surplus** and **producer surplus** to address that question. We will begin with consumer surplus, which is a measure of how much net benefit buyers enjoy from the ability to participate in a particular market.

To get an intuitive feel for this concept, consider the last thing you purchased. Maybe it was a cup of coffee, a new pair of shoes, or a new car. Whatever it was, think of how much you actually paid for it. Now contrast that price with the maximum amount you would have been *willing to pay* for it instead of going without it altogether. If those two numbers are different, we say you received some consumer surplus from your purchase. You received a “bargain” because you were willing to pay more than you had to pay.

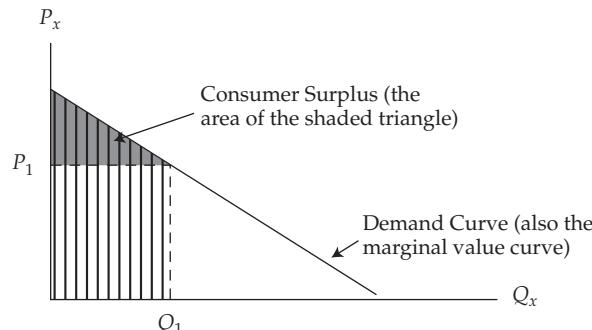
Earlier we referred to the law of demand, which says that as price falls, consumers are willing to buy more of the good. This observation translates into a negatively sloped demand curve. Alternatively, we could say that the highest price that consumers are willing to pay for an additional unit declines as they consume more and more of it. In this way, we can interpret their *willingness to pay* as a measure of how much they *value* each additional unit of the good. This point is very important: To purchase a unit of some good, consumers must give up something else they value. So the price they are willing to pay for an additional unit of a good is a measure of how much they value that unit, in terms of the other goods they must sacrifice to consume it.

If demand curves are negatively sloped, it must be because the value of each additional unit of the good falls the more of it they consume. We will explore this concept further later, but for now it is enough to recognize that the demand curve can thus be considered a **marginal value curve** because it shows the highest price consumers are willing to pay for each *additional* unit. In effect, the demand curve is the willingness of consumers to pay for each additional unit.

This interpretation of the demand curve allows us to measure the total value of consuming any given quantity of a good: It is the sum of all the marginal values of each unit consumed, up to and including the last unit. Graphically, this measure translates into the area under the consumer’s demand curve, up to and including the last unit consumed, as shown in Exhibit 12, in which the consumer is choosing to buy Q_1 units of the good at a price of P_1 . The **marginal value** of the Q_1^{th} unit is clearly P_1 , because that is the highest price the consumer is willing to pay for that unit. Importantly, however, the marginal value of each unit *up to* the Q_1^{th} is greater than P_1 .⁷

Because the consumer would have been willing to pay more for each of those units than she actually paid (P_1), then we can say she received more value than the cost to her of buying them. This concept is referred to as **consumer surplus**, and it is defined as the difference between the value that the consumer places on those units and the amount of money that was required to pay for them. The total value of Q_1 is thus the area of the vertically crosshatched trapezoid in Exhibit 12. The **total expenditure** is only the area of the rectangle with height P_1 and base Q_1 . The total consumer surplus received from buying Q_1 units at a level price of P_1 per unit is the *difference* between the area under the demand curve, on the one hand, and the area of the rectangle, $P_1 \times Q_1$, on the other hand. That area is shown as the lightly shaded triangle.

⁷ This assumes that all units of the good are sold at the same price P_1 . Because the demand curve is negatively sloped, all units up to the Q_1^{th} have marginal values greater than that price.

Exhibit 12 Consumer Surplus

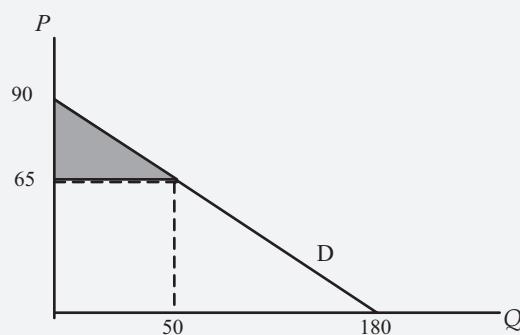
Note: Consumer surplus is the area beneath the demand curve and above the price paid.

EXAMPLE 9**Calculating Consumer Surplus**

A market demand function is given by the equation $Q^d = 180 - 2P$. Determine the value of consumer surplus if price is equal to 65.

Solution:

First, insert 65 into the demand function to find the quantity demanded at that price: $Q^d = 180 - 2(65) = 50$. Then, to make drawing the demand curve easier, invert the demand function by solving it for P in terms of Q : $P = 90 - 0.5Q$. Note that the price intercept is 90, and the quantity intercept is 180. Draw the demand curve:



Find the area of the triangle above the price and below the demand curve, up to quantity 50: Area of a triangle is given as $1/2 \times \text{Base} \times \text{Height} = (1/2)(50)(25) = 625$.

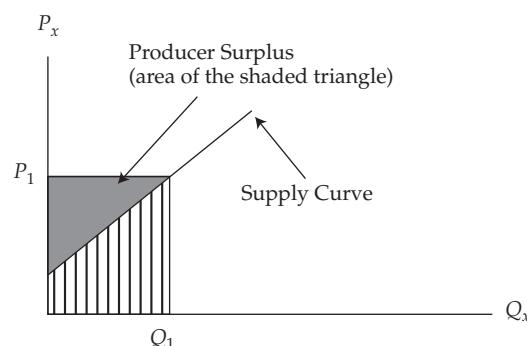
3.10 Producer Surplus—Revenue minus Variable Cost

In this section, we discuss a concept analogous to consumer surplus called **producer surplus**. It is the difference between the total revenue sellers receive from selling a given amount of a good, on the one hand, and the total variable cost of producing that amount, on the other hand. **Variable costs** are those costs that change when the level of output changes. Total revenue is simply the total quantity sold multiplied by the price per unit.

The total variable cost (variable cost per unit times units produced) is measured by the area beneath the supply curve, and it is a little more complicated to understand. Recall that the supply curve represents the lowest price that sellers would be willing to accept for each additional unit of a good. In general, that amount is the cost of producing that next unit, called **marginal cost**. Clearly, a seller would never intend to sell a unit of a good for a price *lower* than its marginal cost, because she would lose money on that unit. Alternatively, a producer should be willing to sell that unit for a price that is *higher* than its marginal cost because it would contribute something toward fixed cost and profit, and obviously the higher the price the better for the seller. Hence, we can interpret the marginal cost curve as the lowest price sellers would accept for each quantity, which basically means the marginal cost curve is the supply curve of any competitive seller. The market supply curve is simply the aggregation of all sellers' individual supply curves, as we showed in section 3.5.

Marginal cost curves are likely to have positive slopes. (It is the logical result of the law of diminishing marginal product, which will be discussed in a later reading.) In Exhibit 13, we see such a supply curve. Because its height is the marginal cost of each additional unit, the total variable cost of Q_1 units is measured as the area beneath the supply curve, up to and including that Q_1^{th} unit, or the area of the vertically cross-hatched trapezoid. But each unit is being sold at the same price P_1 , so total revenue to sellers is the rectangle whose height is P_1 and base is total quantity Q_1 . Because sellers would have been willing to accept the amount of money represented by the trapezoid, but they actually received the larger area of the rectangle, we say they received producer surplus equal to the area of the shaded triangle. So sellers also got a "bargain" because they received a higher price than they would have been willing to accept for each unit.

Exhibit 13 Producer Surplus



Note: Producer surplus is the area beneath the price and above the supply curve.

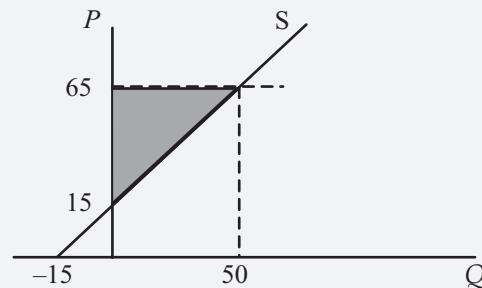
EXAMPLE 10

Calculating the Amount of Producer Surplus

A market supply function is given by the equation $Q^s = -15 + P$. Determine the value of producer surplus if price were equal to 65.

Solution:

First, insert 65 into the supply function to find quantity supplied at that price: $Q^s = -15 + (65) = 50$. Then, to make drawing the supply curve easier, invert the supply function by solving for P in terms of Q : $P = 15 + Q$. Note that the price intercept is 15, and the quantity intercept is -15 . Draw the supply curve:



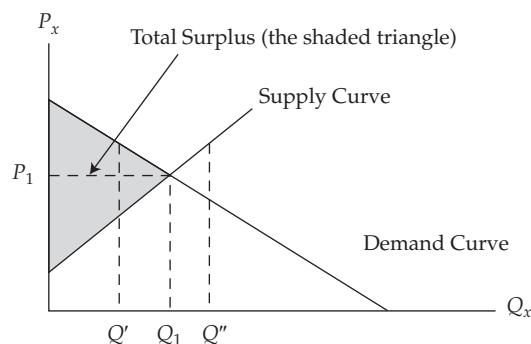
Find the area of the triangle below the price, above the supply curve, up to a quantity of 50: Area = $1/2$ Base \times Height = $(1/2)(50)(50) = 1,250$.

3.11 Total Surplus—Total Value minus Total Variable Cost

In the previous sections, we have seen that consumers and producers both receive “a bargain” when they are allowed to engage in a mutually beneficial, voluntary exchange with one another. For every unit up to the equilibrium unit traded, buyers would have been willing to pay more than they actually had to pay. Additionally, for every one of those units, sellers would have been willing to sell it for less than they actually received. The total value to buyers was greater than the total variable cost to sellers. The difference between those two values is called **total surplus**, and it is made up of the sum of consumer surplus and producer surplus. Note that the way the total surplus is divided between consumers and producers depends on the steepness of the demand and supply curves. If the supply curve is steeper than the demand curve, more of the surplus is being captured by producers. If the demand curve is steeper, consumers capture more of the surplus.

In a fundamental sense, total surplus is a measure of society’s gain from the voluntary exchange of goods and services. Whenever total surplus increases, society gains. An important result of market equilibrium is that total surplus is maximized at the equilibrium price and quantity. Exhibit 14 combines the supply curve and the demand curve to show market equilibrium and total surplus, represented as the area of the shaded triangle. The area of that triangle is the difference between the trapezoid of total value to society’s buyers and the trapezoid of total resource cost to society’s sellers. If price measures dollars (or euros) per unit, and quantity measures units per month, then the measure of total surplus is dollars (euros) per month. It is the “bargain” that buyers and sellers together experience when they voluntarily trade the good in a market. If the market ceased to exist, that would be the monetary value of the loss to society.

Exhibit 14 Total Surplus as the Area beneath the Demand Curve and above the Supply Curve



3.12 Markets Maximize Society's Total Surplus

Recall that the market demand curve can be considered the willingness of consumers to pay for each additional unit of a good. Hence, it is society's marginal value curve for that good. Additionally, the market supply curve represents the marginal cost to society to produce each additional unit of that good, assuming no positive or negative externalities. (An **externality** is a case in which production costs or the consumption benefits of a good or service spill over onto those who are not producing or consuming the good or service; a spillover cost (e.g., pollution) is called a **negative externality**, a spillover benefit (e.g., literacy programs) is called a **positive externality**.)

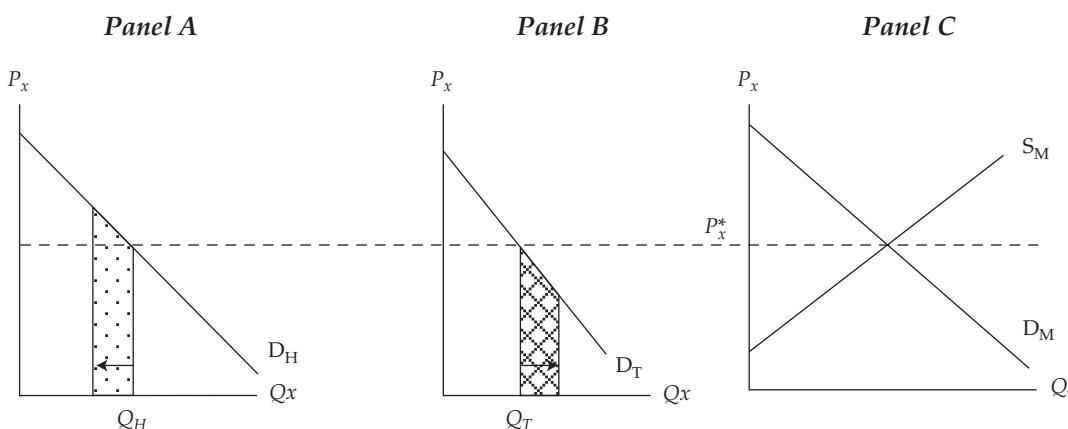
At equilibrium, where demand and supply curves intersect, the highest price that someone is willing to pay is just equal to the lowest price that a seller is willing to accept, which is the marginal cost of that unit of the good. In Exhibit 14, that equilibrium quantity is Q_1 . Now, suppose that some influence on the market caused less than Q_1 units to be traded, say only Q' units. Note that the marginal value of the Q'^{th} unit exceeds society's marginal cost to produce it. In a fundamental sense, we could say that society *should* produce and consume it, as well as the next, and the next, all the way up to Q_1 . Or suppose that some influence caused more than Q_1 to be produced, say Q'' units. Then what can we say? For all those units beyond Q_1 , and up to Q'' , society incurred greater cost than the value it received from consuming them. We could say that society *should not* have produced and consumed those additional units. Total surplus was reduced by those additional units because they cost more in the form of resources than the value they provided for society when they were consumed.

There is reason to believe that markets usually trend toward equilibrium and that the condition of equilibrium itself is also optimal in a welfare sense. To delve a little more deeply, consider two consumers, Helen Smith and Tom Warren, who have access to a market for some good, perhaps gasoline or shoes or any other consumption good. We could depict their situations using their individual demand curves juxtaposed on an exhibit of the overall market equilibrium, as in Exhibit 15 where Smith's and Warren's individual demands for a particular good are depicted along with the market demand and supply of that same good. (The horizontal axes are scaled differently because the market quantity is so much greater than either consumer's quantity, but the price axes are identical.)

At the market price of P_x^* , Smith chooses to purchase Q_H , and Warren chooses to purchase Q_T because at that price, the marginal value for each of the two consumers is just equal to the price they have to pay per unit. Now, suppose someone removed

one unit of the good from Smith and presented it to Warren. In Panel A of Exhibit 15, the loss of value experienced by Smith is depicted by the dotted trapezoid, and in Panel B of Exhibit 15, the gain in value experienced by Warren is depicted by the crosshatched trapezoid. Note that the increase in Warren's value is necessarily less than the loss in Smith's. Recall that consumer surplus is value minus expenditure. Total consumer surplus is reduced when individuals consume quantities that do not yield equal marginal value to each one. Conversely, when all consumers face the identical price, they will purchase quantities that equate their marginal values across all consumers. Importantly, that behavior maximizes total consumer surplus.

Exhibit 15 How Total Surplus Can Be Reduced by Rearranging Quantity

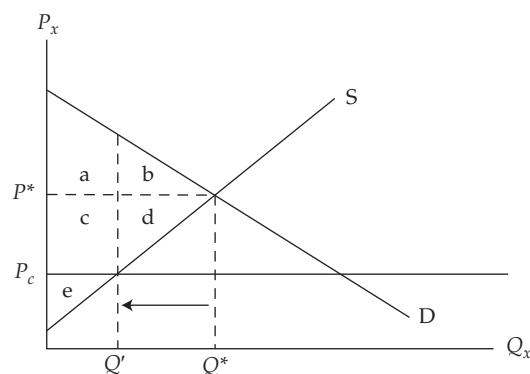


Note: Beginning at a competitive market equilibrium, when one unit is taken from Smith and presented to Warren, total surplus is reduced.

A precisely analogous argument can be made to show that when all producers produce quantities such that their marginal costs are equated across all firms, total producer surplus is maximized. The result of this analysis is that when all consumers face the same market equilibrium price and are allowed to buy all they desire at that price, and when all firms face that same price and are allowed to sell as much as they want at that price, the total of consumer and producer surplus (total surplus) is maximized from that market. This result is the beauty of free markets: They maximize society's net benefit from production and consumption of goods and services.

3.13 Market Interference: The Negative Impact on Total Surplus

Sometimes, lawmakers determine that the market price is “too high” for consumers to pay, so they use their power to impose a ceiling on price below the market equilibrium price. Some examples of ceilings include rent controls (limits on increases in the rent paid for apartments), limits on the prices of medicines, and laws against “price gouging” after a hurricane (i.e., charging opportunistically high prices for goods such as bottled water or plywood). Certainly, price limits benefit anyone who had been paying the old higher price and can still buy all they want at the lower ceiling price. However, the story is more complicated than that. Exhibit 16 shows a market in which a ceiling price, P_c , has been imposed below equilibrium. Let's examine the full impact of such a law.

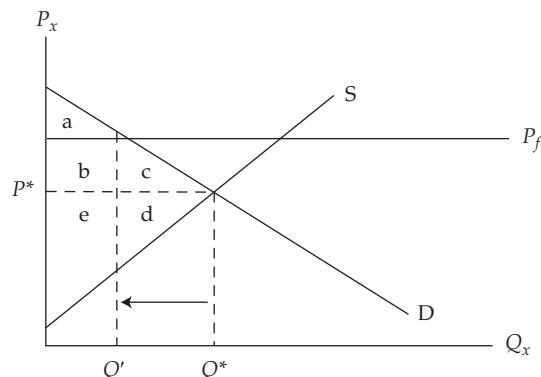
Exhibit 16 A Price Ceiling

Note: A price ceiling transfers surplus equal to area c from sellers to buyers, but it destroys surplus equal to area b + d, called a deadweight loss.

Prior to imposition of the ceiling price, equilibrium occurs at (P^*, Q^*) , and total surplus equals the area given by $a + b + c + d + e$. It consists of consumer surplus given by $a + b$, and producer surplus given by $c + d + e$. When the ceiling is imposed, two things happen: Buyers would like to purchase more at the lower price, but sellers are willing now to sell less. Regardless of how much buyers would like to purchase, though, only Q' would be offered for sale. Clearly, the total quantity that actually gets traded has fallen, and this has some serious consequences. For one thing, any buyer who is still able to buy the Q' quantity has clearly been given a benefit. They used to pay P^* and now pay only P_c per unit. Those buyers gain consumer surplus equal to rectangle c , which used to be part of seller surplus. Rectangle c is surplus that has been transferred from sellers to buyers, but it still exists as part of total surplus. Disturbingly, though, there is a loss of consumer surplus equal to triangle b and a loss of producer surplus equal to triangle d . Those measures of surplus simply no longer exist at the lower quantity. Clearly, surplus cannot be enjoyed on units that are neither produced nor consumed, so that loss of surplus is called a **deadweight loss** because it is surplus that is lost by one or the other group but not transferred to anyone. Thus, after the imposition of a price ceiling at P_c , consumer surplus is given by $a + c$, producer surplus by e , and the deadweight loss is $b + d$.⁸

Another example of price interference is a **price floor**, in which lawmakers make it illegal to buy or sell a good or service below a certain price, which is above equilibrium. Again, some sellers who are still able to sell at the now higher floor price benefit from the law, but that's not the whole story. Exhibit 17 shows such a floor price, imposed at P_f above free market equilibrium.

⁸ Technically, the statement assumes that the limited sales are allocated to the consumers with the highest valuations. A detailed explanation, however, is outside the scope of this reading.

Exhibit 17 A Price Floor

Note: A price floor transfers surplus equal to area b from buyers to sellers, but it destroys surplus equal to area c + d, called a deadweight loss.

At free market equilibrium quantity Q^* , total surplus is equal to $a + b + c + d + e$, consisting of consumer surplus equal to area $a + b + c$, and producer surplus equal to area $e + d$. When the floor is imposed, sellers would like to sell more, but buyers would choose to purchase less. Regardless of how much producers want to sell, however, only Q' will be purchased at the new higher floor price. Those sellers who can still sell at the higher price benefit at the expense of the buyers: There is a transfer of surplus from buyers to sellers equal to rectangle b. Regrettably, however, that's not all. Buyers also lose consumer surplus equal to triangle c, and sellers lose producer surplus equal to triangle d.⁹ Once again, no one can benefit from units that are neither produced nor consumed, so there is a deadweight loss equal to triangle c plus triangle d. As a result of the floor, the buyer's surplus is reduced to triangle a.

A good example of a price floor is the imposition of a legal minimum wage in the United States, the United Kingdom, and many other countries. Although controversy remains among some economists on the empirical effects of the minimum wage, most economists continue to believe that a minimum wage can reduce employment. Although some workers will benefit, because they continue to work at the higher wage, others will be harmed because they will no longer be working at the increased wage rate.

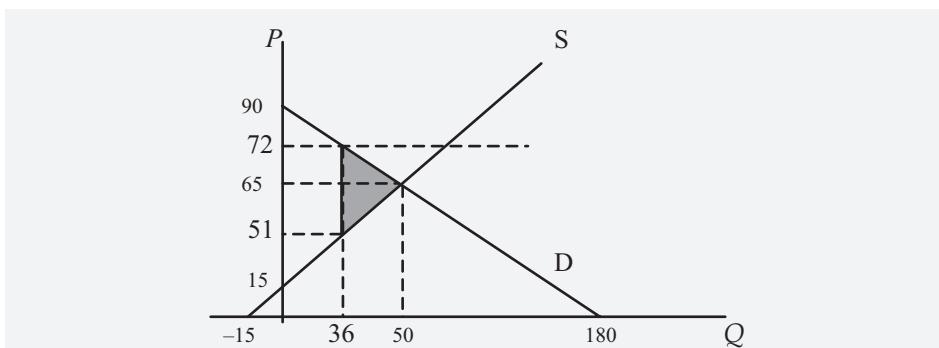
EXAMPLE 11**Calculating the Amount of Deadweight Loss from a Price Floor**

A market has demand function given by the equation $Q^d = 180 - 2P$, and supply function given by the equation $Q^s = -15 + P$. Calculate the amount of deadweight loss that would result from a price floor imposed at a level of 72.

Solution:

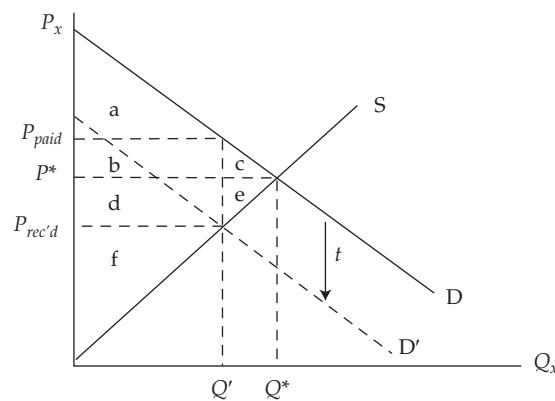
First, solve for equilibrium price of 65 and quantity 50. Then, invert the demand function to find $P = 90 - 0.5Q$, and the supply function to find $P = 15 + Q$. Use these functions to draw the demand and supply curves:

⁹ Technically, this statement assumes that sales are made by the lowest cost producers. A discussion of the point is outside the scope of this reading.



Insert the floor price of 72 into the demand function to find that only 36 would be demanded at that price. Insert 36 into the supply function to find the price of 51 that corresponds to a quantity of 36. Because the price floor would reduce quantity from its equilibrium value of 50 to the new value of 36, the deadweight loss would occur because those 14 units are not now being produced and consumed under the price floor. So deadweight loss equals the area of the shaded triangle: $1/2 \text{ Base} \times \text{Height} = (1/2)(72 - 51)(50 - 36) = 147$.

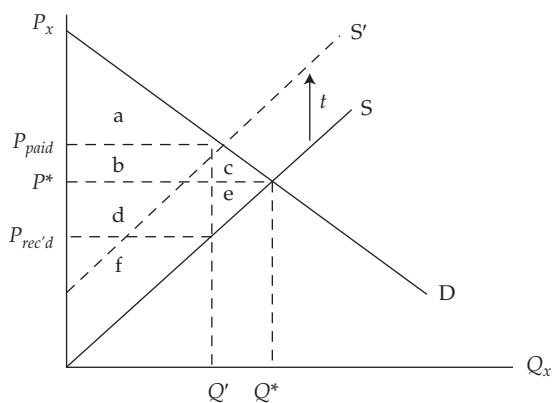
Still other policies can interfere with the ability of prices to allocate society's resources. Governments do have legitimate functions to perform in society, and they need to have revenue to finance them. So they often raise revenue by imposing taxes on various goods or activities. One such policy is a per-unit tax, such as an excise tax. By law, this tax could be imposed either on buyers or on sellers, but we shall see that it really doesn't matter at all who legally must pay the tax, the result is the same: more deadweight loss. Exhibit 18 depicts such a tax imposed in this case on buyers. Here, the law simply says that whenever a buyer purchases a unit of some good, he or she must pay a tax of some amount t per unit. Recall that the demand curve is the highest price willingly paid for each quantity. Because buyers probably do not really care who receives the money, government or the seller, their gross willingness to pay is still the same. Because they must pay t dollars to the government, however, their net demand curve would shift vertically downward by t per unit. Exhibit 18 shows the result of such a shift.

Exhibit 18 A Per-Unit Tax on Buyers


Note: A tax on buyers shifts the demand curve downward by t , imposing a burden on both buyers and sellers, shifting some of the surplus to government but leaving a deadweight loss equal to c plus e .

Originally, the pre-tax equilibrium is where D and S intersect at (P^*, Q^*) . Consumer surplus is given by triangle a plus rectangle b plus triangle c , and producer surplus consists of triangle f plus rectangle d plus triangle e . When the tax is imposed, the demand curve shifts vertically downward by the tax per unit, t . This shift results in a new equilibrium at the intersection of S and D' . That new equilibrium price is received by sellers ($P_{rec'd}$). However, buyers now must pay an additional t per unit to government, resulting in a total price paid (P_{paid}) that is higher than before. Sellers receive a lower price and buyers pay a higher price than pretax, so both suffer a burden as a result of this tax, even though it was legally imposed only on buyers. Buyers now have consumer surplus that has been reduced by rectangle b plus triangle c ; thus, post-tax consumer surplus is $(a + b + c) - (b + c) = a$. Sellers now have producer surplus that has been reduced by rectangle d plus triangle e ; thus post-tax producer surplus is $(f + d + e) - (d + e) = f$. Government receives tax revenue of t per unit multiplied by Q' units. Its total revenue is rectangle b plus rectangle d . Note that the total loss to buyers and sellers ($b + c + d + e$) is greater than the revenue transferred to government ($b + d$), so that the tax resulted in a deadweight loss equal to triangle c plus triangle e as $(b + c + d + e) - (b + d) = c + e$.

How would things change if the tax had legally been imposed on sellers instead of buyers? To see the answer, note that the supply curve is the lowest price willingly accepted by sellers, which is their marginal cost. If they now must pay an additional t dollars per unit to government, their lowest acceptable price for each unit is now higher. We show this by shifting the supply curve vertically upward by t dollars per unit, as shown in Exhibit 19.

Exhibit 19 A Per-Unit Tax on Sellers

Note: A tax on sellers shifts the supply curve upward by t . Everything is exactly the same as in the case of imposing the tax on buyers.

The new equilibrium occurs at the intersection of S' and D , resulting in the new equilibrium price paid by buyers, P_{paid} . Sellers are paid this price but must remit t dollars per unit to the government, resulting in an after-tax price received ($P_{rec'd}$) that is lower than before the tax. In terms of overall result, absolutely nothing is different from the case in which buyers had the legal responsibility to pay the tax. Tax revenue to the government is the same, buyers' and sellers' reduction in surplus is identical to the previous case, and the deadweight loss is the same as well.

Notice that the share of the total burden of the tax need not be equal for buyers and sellers. In our example, sellers experienced a greater burden than buyers did, regardless of who had the legal responsibility to pay the tax. The relative burden from a tax falls disproportionately on the group (buyers or sellers) that has the steeper curve. In our example, the demand curve is flatter than the supply curve (just slightly so), so buyers bore proportionately less of the burden. Just the reverse would be true if the demand curve had been steeper than the supply curve.

All of the policies we have examined involve government interfering with free markets. Other examples include imposing tariffs on imported goods, setting quotas on imports, or banning the trade of goods. Additionally, governments often impose regulations on the production or consumption of goods to limit or correct the negative effects on third parties that cannot be captured in free market prices. Even the most ardent of free market enthusiasts recognize the justification of some government intervention in the case of public goods, such as for national defense, or where prices do not reflect true marginal social value or cost, as in externalities such as pollution. Social considerations can trump pure economic efficiency, as in the case of child labor laws or human trafficking. What does come from the analysis of markets, however, is the recognition that when social marginal benefits are truly reflected in market demand curves and social marginal costs are truly reflected in supply curves, total surplus is maximized when markets are allowed to operate freely. Moreover, when society does choose to impose legal restrictions, market analysis of the kind we have just examined provides society with a means of at least assessing the deadweight losses that such policies extract from total surplus. In that way, policy makers can perform logical, rigorous cost benefit assessments of their proposed policies to inform their decisions.

EXAMPLE 12**Calculating the Effects of a Per-Unit Tax on Sellers**

A market has a demand function given by the equation $Q^d = 180 - 2P$, and a supply function given by the equation $Q^s = -15 + P$, where price is measured in euros per unit. A tax of €2 per unit is imposed on sellers in this market.

- 1 Calculate the effect on the price paid by buyers and the price received by sellers.
- 2 Demonstrate that the effect would be unchanged if the tax had been imposed on buyers instead of sellers.

Solution to 1:

Determine the pre-tax equilibrium price and quantity by equating supply and demand: $180 - 2P = -15 + P$. Therefore $P^* = €65$ before tax. If the tax is imposed on sellers, the supply curve will shift upward by €2. So, to begin, we need to invert the supply function and the demand function: $P = 15 + Q^s$ and $P = 90 - 0.5Q^d$. Now, impose the tax on sellers by increasing the value of P by €2 at each quantity. This step simply means increasing the price intercept by €2. Because sellers must pay €2 tax per unit, the lowest price they are willing to accept for each quantity rises by that amount: $P' = 17 + Q^s$, where “ P prime” indicates the new function after imposition of the tax. Because the tax was not imposed on buyers, the inverse demand function remains as it was. Solve for the new equilibrium price and quantity: $90 - 0.5Q = 17 + Q$, so new after-tax $Q = 48.667$. By inserting that quantity into the new inverse demand function, we find that $P_{paid} = €65.667$. This amount is paid by buyers to sellers, but because sellers are responsible for paying the €2 tax, they receive only $€65.667 - €2 = €63.667$, after tax. So we find that the tax on sellers has increased the price to buyers by €0.667 while reducing the price received by sellers by €1.33. Out of the €2 tax, buyers bear one-third of the burden and sellers bear two-thirds of the burden. This result is because the demand curve is half as steep as the supply curve. The group with the steepest, less elastic, curve bears the greater burden of a tax, regardless of on whom the legal incidence of the tax is imposed.

Solution to 2:

Instead of adding €2 to the price intercept of supply curve, we now subtract €2 from the price intercept of the demand curve. This step is because buyers’ willingness to pay sellers has been reduced by the €2 they must pay in tax per unit. Buyers really don’t care who receives their money, they are interested only in the greatest amount they are willing to pay for each quantity. So the new inverse demand function is: $P'' = 88 - 0.5Q$. Using this new inverse demand, we now solve for equilibrium: $88 - 0.5Q = 15 + Q$. (Because buyers must pay the tax, we leave the old supply curve unchanged.) The new equilibrium quantity is therefore $Q = 48.667$, which is exactly as it was when sellers had the obligation to pay the tax. Inserting that number into the old supply function gives us the new equilibrium price of €63.667, which is what buyers must pay sellers. Recall, however, that now buyers must pay €2 in tax per unit, so the price buyers pay after tax is $€63.667 + €2 = €65.667$. So nothing changes when we impose the statutory obligation on buyers instead of sellers. They still share the ultimate burden of the tax in exactly the same proportion as when sellers had to send the €2 to the taxing authority.

We have seen that government interferences, such as price ceilings, price floors, and taxes, result in imbalances between demand and supply. In general, anything else that intervenes in the process of buyers and sellers finding the equilibrium price can cause imbalances as well. In the simple model of demand and supply, it is assumed that buyers and sellers can interact without cost. Often, however, there can be costs associated with finding a buyer's or a seller's counterpart. There could be a buyer who is willing to pay a price higher than some seller's lowest acceptable price, but if the two cannot find one another, there will be no transaction, resulting in a deadweight loss. The costs of matching buyers with sellers are generally referred to as **search costs**, and they arise because of frictions inherent in the matching process. When these costs are significant, an opportunity may arise for a third party to provide a valuable service by reducing those costs. This role is played by brokers. Brokers do not actually become owners of a good or service that is being bought, but they serve the role of locating buyers for sellers or sellers for buyers. (Dealers, however, actually take possession of the item in anticipation of selling it to a future buyer.) To the extent that brokers serve to reduce search costs, they provide value in the transaction, and for that value they are able to charge a brokerage fee. Although the brokerage fee could certainly be viewed as a transactions cost, it is really a price charged for the service of reducing search costs. In effect, any impediment in the dissemination of information about buyers' and sellers' willingness to exchange goods can cause an imbalance in demand and supply. So anything that improves that information flow can add value. In that sense, advertising can add value to the extent that it informs potential buyers of the availability of goods and services.

DEMAND ELASTICITIES

4

The general model of demand and supply can be highly useful in understanding directional changes in prices and quantities that result from shifts in one or the other curve. At a deeper quantitative level, though, we often need to measure just *how* sensitive quantity demanded or supplied is to changes in the independent variables that affect them. Here is where the concept of *elasticity of demand and supply* plays a crucial role in microeconomics. We will examine several elasticities of demand, but the crucial element is that fundamentally all elasticities are calculated the same way: they are ratios of percentage changes. Let us begin with the sensitivity of quantity demanded to changes in the own-price.

4.1 Own-Price Elasticity of Demand

Recall that when we introduced the concept of a demand function with Equation 1 earlier, we were simply theorizing that quantity demanded of some good, such as gasoline, is dependent on several other variables, one of which is the price of gasoline itself. We referred to the law of demand that simply states the inverse relationship between the quantity demanded and the price. Although that observation is useful, we might want to dig a little deeper and ask, Just how sensitive is quantity demanded to changes in the price of gasoline? Is it highly sensitive, so that a very small rise in price is associated with an enormous fall in quantity, or is the sensitivity only minimal? It might be helpful if we had a convenient measure of this sensitivity.

In Equation 3, we introduced a hypothetical household demand function for gasoline, assuming that the household's income and the price of another good (automobiles) were held constant. It supposedly described the purchasing behavior of a household regarding its demand for gasoline. That function was given by the simple

linear expression $Q_x^d = 11.2 - 0.4P_x$. If we were to ask how sensitive quantity is to changes in price in that expression, one plausible answer would be simply to recognize that, according to that demand function, whenever price changes by one unit, quantity changes by 0.4 units in the opposite direction. That is to say, if price were to rise by \$1, quantity would fall by 0.4 gallons per week, so the coefficient on the price variable (-0.4) could be the measure of sensitivity we are seeking.

There is a fundamental drawback, however, associated with that measure. Notice that the -0.4 is measured in gallons of gasoline per dollar of price. It is crucially dependent on the *units* in which we measured Q and P . If we had measured the price of gasoline in cents per gallon, instead of dollars per gallon, then the exact same household behavior would be described by the alternative equation $Q_x^d = 11.2 - 0.004P_x$. So, although we could choose the coefficient on price as our measure of sensitivity, we would always need to recall the units in which Q and P were measured when we wanted to describe the sensitivity of gasoline demand. That could be cumbersome.

Because of this drawback, economists prefer to use a gauge of sensitivity that does not depend on units of measure. That metric is called **elasticity**, and it is defined as the ratio of *percentage changes*. It is a general measure of how sensitive one variable is to any other variable. For example, if some variable y depends on some other variable x in the following function: $y = f(x)$, then the elasticity of y with respect to x is defined to be the percentage change in y divided by the percentage change in x , or $\%Δy/\%Δx$. In the case of **own-price elasticity of demand**, that measure is¹⁰

$$E_{P_x}^d = \frac{\%ΔQ_x^d}{\%ΔP_x} \quad (23)$$

Notice that this measure is independent of the units in which quantity and price are measured. If, for example, when price rises by 10 percent, quantity demanded falls by 8 percent, then elasticity of demand is simply -0.8 . It does not matter whether we are measuring quantity in gallons per week or liters per day, and it does not matter whether we measure price in dollars per gallon or euros per liter; 10 percent is 10 percent, and 8 percent is 8 percent. So the ratio of the first to the second is still -0.8 .

We can expand Equation 23 algebraically by noting that the percentage change in any variable x is simply the change in x (denoted “ $Δx$ ”) divided by the level of x . So, we can rewrite Equation 23, using a couple of simple steps, as

$$E_{P_x}^d = \frac{\%ΔQ_x^d}{\%ΔP_x} = \frac{\frac{\Delta Q_x^d}{Q_x^d}}{\frac{\Delta P_x}{P_x}} = \left(\frac{\Delta Q_x^d}{\Delta P_x} \right) \left(\frac{P_x}{Q_x^d} \right) \quad (24)$$

To get a better idea of price elasticity, it might be helpful to use our hypothetical market demand function: $Q_x^d = 11,200 - 400P_x$. For linear demand functions, the first term in the last line of Equation 24 is simply the slope coefficient on P_x in the demand function, or -400 . (Technically, this term is the first derivative of Q_x^d with respect to P_x , dQ_x^d/dP_x , which is the slope coefficient for a linear demand function.) So, the elasticity of demand in this case is -400 multiplied by the ratio of price to quantity. Clearly in this case, we need to choose a price at which to calculate the elasticity coefficient. Let's choose the original equilibrium price of \$3. Now, we need to find the quantity associated with that particular price by inserting 3 into the demand function and finding $Q = 10,000$. The result of our calculation is that at a price of 3, the elasticity of our market demand function is $-400 (3/10,000) = -0.12$. How do we

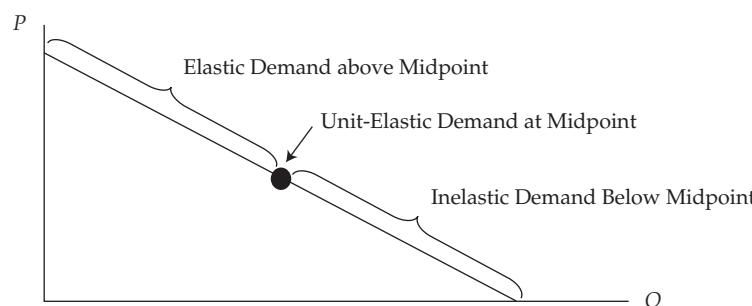
¹⁰ The reader will also encounter the Greek letter epsilon ($ε$) being used in the notation for elasticities.

interpret that value? It means, simply, that when price equals 3, a 1 percent rise in price would result in a fall in quantity demanded of only 0.12 percent. (You should try calculating price elasticity when price is equal to, say, \$4. Do you find that elasticity equals -0.167 ?¹¹)

In our particular example, when price is \$3 per gallon, demand is not very sensitive to changes in price, because a 1 percent rise in price would reduce quantity demanded by only 0.12 percent. Actually, that is not too different from empirical estimates of the actual demand elasticity for gasoline in the United States. When demand is not very sensitive to price, we say demand is **inelastic**. To be precise, when the *magnitude* (ignoring algebraic sign) of the own-price elasticity coefficient has a value less than one, demand is defined to be inelastic. When that magnitude is greater than one, demand is defined to be **elastic**. And when the elasticity coefficient is equal to negative one, demand is said to be **unit elastic**, or **unitary elastic**. Note that if the law of demand holds, own-price elasticity of demand will always be negative, because a rise in price will be associated with a fall in quantity demanded, but it can be either elastic or inelastic. In our hypothetical example, suppose the price of gasoline was very high, say \$15 per gallon. In this case, the elasticity coefficient would be -1.154 . Therefore, because the magnitude of the elasticity coefficient is greater than one, we would say that demand is elastic at that price.¹¹

By examining Equation 24, we should be able to see that for a linear demand curve the elasticity depends on where we calculate it. Note that the first term, $\Delta Q/\Delta P$, will remain constant along the entire demand curve because it is simply the inverse of the slope of the demand curve. But the second term, P/Q , clearly changes depending on where we look. At very low prices, P/Q is very small, so demand is inelastic. But at very high prices, Q is low and P is high, so the ratio P/Q is very high, and demand is elastic. Exhibit 20 illustrates a characteristic of all negatively sloped linear demand curves. Above the midpoint of the curve, demand is elastic; below the midpoint, demand is inelastic; and at the midpoint, demand is unit elastic.

Exhibit 20 The Elasticity of a Linear Demand Curve



Note: For all negatively sloped, linear demand curves, elasticity varies depending on where it is calculated.

Sometimes, we might not have the entire demand function or demand curve, but we might have just two observations on price and quantity. In this case, we do not know the slope of the demand curve at a given point because we really cannot say that it is even a linear function. For example, suppose we know that when price is 5,

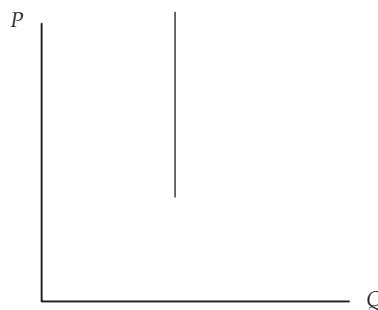
¹¹ For evidence on price elasticities of demand for gasoline, see Molly Espey, "Explaining the Variation in Elasticity Estimates of Gasoline Demand in the United States: A Meta-analysis," *Energy Journal*, vol. 17, no. 3(1996): 49–60. The robust estimates were about -0.26 for short-run elasticity (less than 1 year) and -0.58 for more than a year.

quantity demanded is 9,200, and when price is 6, quantity demanded is 8,800, but we do not know anything more about the demand function. Under these circumstances, economists use something called **arc elasticity**. Arc elasticity of demand is still defined as the percentage change in quantity demanded divided by the percentage change in price. However, because the choice of base for calculating percentage changes has an effect on the calculation, economists have chosen to use the *average* quantity and the *average* price as the base for calculating the percentage changes. (Suppose, for example, that you are making a wage of €10 when your boss says, “I’ll increase your wage by 10 percent.” You are then earning €11. But later that day, if your boss then reduces your wage by 10 percent, you are then earning €9.90. So, by receiving first a 10 percent raise and then a 10 percent cut in wage, you are worse off. The reason for this is that we typically use the original value as the base, or denominator, for calculating percentages.) In our example, then, the arc elasticity of demand would be:

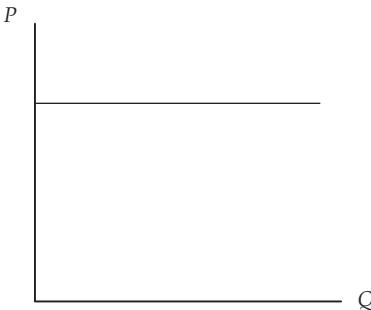
$$E = \frac{\frac{\Delta Q}{Q_{avg}}}{\frac{\Delta P}{P_{avg}}} = \frac{\frac{-400}{9,000}}{\frac{1}{5.5}} = -0.244$$

There are two special cases in which linear demand curves have the same elasticity at all points: vertical demand curves and horizontal demand curves. Consider a vertical demand curve, as in Exhibit 21 Panel A, and a horizontal demand curve, as in Panel B. In the first case, the quantity demanded is the same, regardless of price. Certainly, there could be no demand curve that is perfectly vertical at *all* possible prices, but over some range of prices it is not unreasonable that the same quantity would be purchased at a slightly higher price or a slightly lower price. Perhaps an individual’s demand for, say, mustard might obey this description. Obviously, in that price range, quantity demanded is not at all sensitive to price and we would say that demand is **perfectly inelastic** in that range.

In the second case, the demand is horizontal at some price. Clearly, for an individual consumer, this situation could not occur because it implies that at even an infinitesimally higher price the consumer would buy nothing, whereas at that particular price, the consumer would buy an indeterminately large amount. This situation is not at all an unreasonable description of the demand curve facing a single seller in a perfectly competitive market, such as the wheat market. At the current market price of wheat, an individual farmer could sell all she has. If, however, she held out for a price above market price, it is reasonable that she would not be able to sell any at all because all other farmers’ wheat is a perfect substitute for hers, so no one would be willing to buy any of hers at a higher price. In this case, we would say that the demand curve facing a perfectly competitive seller is **perfectly elastic**.

Exhibit 21 The Extremes of Price Elasticity
Panel A

Note: A vertical demand has zero elasticity and is called perfectly inelastic.

Panel B

Note: A horizontal demand has infinite elasticity and is called perfectly elastic.

Own-price elasticity of demand is our measure of how sensitive the quantity demanded is to changes in the price of a good or service, but what characteristics of a good or its market might be informative in determining whether demand is highly elastic or not? Perhaps the most important characteristic is whether there are close substitutes for the good in question. If there are close substitutes for the good, then if its price rises even slightly, a consumer would tend to purchase much less of this good and switch to the substitute, which is now relatively less costly. If there simply are no substitutes, however, then it is likely that the demand is much less elastic. To understand this more fully, consider a consumer's demand for some broadly defined product such as bread. There really are no close substitutes for the category bread, which includes all types from French bread to pita bread to tortillas and so on. So, if the price of all bread were to rise, perhaps a consumer would purchase a little less of it each week, but probably not a significantly smaller amount. Now, however, consider that the consumer's demand for a particular baker's specialty bread instead of the category "bread" as a whole. Surely, there are close substitutes for Baker Bob's Whole Wheat Bread with Sesame Seeds than for bread in general. We would expect, then, that the demand for Baker Bob's special loaf is much more elastic than for the entire category of bread. This fact is why the demand faced by an individual wheat farmer is much more elastic than the entire market demand for wheat; there are much closer substitutes for *her* wheat than for wheat *in general*.

In finance, there exists the question of whether the demand for common stock is perfectly elastic. That is, are there perfect substitutes for a firm's common shares? If so, then the demand curve for its shares should be perfectly horizontal. If not, then one would expect a negatively sloped demand for shares. If demand is horizontal, then an increase in demand (owing to some influence other than positive new information regarding the firm's outlook) would not increase the share price. In contrast, a purely "mechanical" increase in demand would be expected to increase the price if the demand were negatively sloped. One study looked at evidence from 31 stocks whose weights on the Toronto Stock Exchange 300 Index were changed, owing purely to fully anticipated technical reasons that apparently had no relationship to new information about those firms.¹² That is, the demand for those shares shifted rightward.

¹² Aditya Kaul, Vikas Mehrotra, and Randall Morck, "Demand Curves for Stocks Do Slope Down: New Evidence from an Index Weights Adjustment," *Journal of Finance*, vol. 55(2 April 2000): 893–912.

The authors found that there was a statistically significant 2.3 percent excess return associated with those shares, a finding consistent with a negatively sloped demand curve for common stock.

In addition to the degree of substitutability, other characteristics tend to be generally predictive of a good's elasticity of demand. These include the portion of the typical budget that is spent on the good, the amount of time that is allowed to respond to the change in price, the extent to which the good is seen as necessary or optional, and so on. In general, if consumers tend to spend a very small portion of their budget on a good, their demand tends to be less elastic than if they spend a very large part of their income. Most people spend only a little on, say, toothpaste each month, so it really doesn't matter whether the price rises 10 percent or not. They would probably still buy about the same amount. If the price of housing were to rise significantly, however, most households would try to find a way to reduce the quantity they buy, at least in the long run.

This example leads to another characteristic regarding price elasticity. For most goods and services, the long-run demand is much more elastic than the short-run demand. The reason is that if price were to change for, say, gasoline, we probably would not be able to respond quickly with a significant reduction in the quantity we consume. In the short run, we tend to be locked into modes of transportation, housing and employment location, and so on. The longer the adjustment time, however, the greater the degree to which a household could adjust to the change in price. Hence, for most goods, long-run elasticity of demand is greater than short-run elasticity. Durable goods, however, tend to behave in the opposite way. If the price of washing machines were to fall, people might react quickly because they have an old machine that they know will need to be replaced fairly soon anyway. So when price falls, they might decide to go ahead and make the purchase. If the price of washing machines were to stay low forever, however, it is unlikely that a typical consumer would buy all that many more machines over a lifetime.

Certainly, whether the good or service is seen to be non-discretionary or discretionary would help determine its sensitivity to a price change. Faced with the same percentage increase in prices, consumers are much more likely to give up their Friday night restaurant meal than they are to cut back significantly on staples in their pantry. The more a good is seen as being necessary, the less elastic its demand is likely to be.

In summary, own-price elasticity of demand is likely to be greater (i.e., more sensitive) for items that have many close substitutes, occupy a large portion of the total budget, are seen to be optional instead of necessary, and have longer adjustment times. Obviously, not all of these characteristics operate in the same direction for all goods, so elasticity is likely to be a complex result of these and other characteristics. In the end, the actual elasticity of demand for a particular good turns out to be an empirical fact that can be learned only from careful observation and often, sophisticated statistical analysis.

4.2 Own-Price Elasticity of Demand: Impact on Total Expenditure

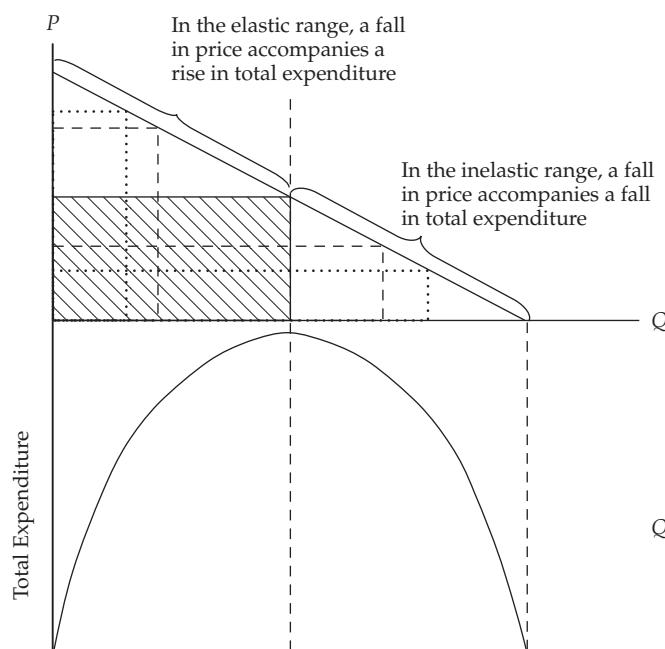
Because of the law of demand, an increase in price is associated with a decrease in the number of units demanded of some good or service. But what can we say about the *total expenditure* on that good? That is, what happens to price times quantity when price falls? Recall that elasticity is defined as the ratio of the percentage change in quantity demanded to the percentage change in price. So if demand is elastic, a decrease in price is associated with a larger percentage rise in quantity demanded. For example, if elasticity were equal to negative two, then the percentage change in quantity demanded would be twice as large as the percentage change in price. It

follows that a 10 percent fall in price would bring about a rise in quantity of greater magnitude, in this case 20 percent. True, each unit of the good has a lower price, but a sufficiently greater number of units are purchased so that total expenditure (price times quantity) would rise as price falls when demand is elastic.

If demand is inelastic, however, a 10 percent fall in price brings about a rise in quantity less than 10 percent in magnitude. Consequently, when demand is inelastic, a fall in price brings about a fall in total expenditure. If elasticity were equal to negative one, (unitary elasticity) the percentage decrease in price is just offset by an equal and opposite percentage increase in quantity demanded, so total expenditure does not change at all.

In summary, when demand is elastic, price and total expenditure move in *opposite* directions. When demand is inelastic, price and total expenditure move in the *same* direction. When demand is unitary elastic, changes in price are associated with *no change* in total expenditure. This relationship is easy to identify in the case of a linear demand curve. Recall from Exhibit 20 that above the midpoint, demand is elastic; and below the midpoint, demand is inelastic. In the upper section of Exhibit 22, total expenditure ($P \times Q$) is measured as the area of a rectangle whose base is Q and height is P . Notice that as price falls, the inscribed rectangles at first grow in size but then become their largest at the midpoint of the demand curve. Thereafter, as price continues to fall, total expenditure falls toward zero. In the lower section of Exhibit 22, total expenditure is shown for each quantity purchased. Note that it reaches a maximum at the quantity that defines the midpoint, or unit-elastic, point on the demand curve.

Exhibit 22 Elasticity and Total Expenditure



Note: Figure depicts the relationship among changes in price, changes in quantity, and changes in total expenditure.

Maximum total expenditure occurs at the unit-elastic point on a linear demand curve (the cross-hatched rectangle).

It should be noted that the relationships just described hold for any demand curve, so it does not matter whether we are dealing with the demand curve of an individual consumer, the demand curve of the market, or the demand curve facing any given

seller. For a market, the total expenditure by buyers becomes the total revenue to sellers in that market. It follows, then, that if market demand is elastic, a fall in price will result in an increase in total revenue to sellers as a whole, and if demand is inelastic, a fall in price will result in a decrease in total revenue to sellers. Clearly, if the demand faced by any given seller were inelastic at the current price, that seller could increase revenue by increasing its price. Moreover, because demand is negatively sloped, the increase in price would decrease total units sold, which would almost certainly decrease total cost. So no one-product seller would ever knowingly choose to set price in the inelastic range of its demand.

4.3 Income Elasticity of Demand: Normal and Inferior Goods

In general, elasticity is simply a measure of how sensitive one variable is to change in the value of another variable. Quantity demanded of a good is a function not only of its own price, but also consumer income. If income changes, the quantity demanded can respond, so the analyst needs to understand the income sensitivity as well as price sensitivity.

Income elasticity of demand is defined as the percentage change in quantity demanded divided by the percentage change in income (I), holding all other things constant, and can be represented as in Equation 25.

$$E_I^d = \frac{\% \Delta Q_x^d}{\% \Delta I} = \frac{\frac{\Delta Q_x^d}{Q_x^d}}{\frac{\Delta I}{I}} = \left(\frac{\Delta Q_x^d}{\Delta I} \right) \left(\frac{I}{Q_x^d} \right) \quad (25)$$

Note that the structure of this expression is identical to the structure of own-price elasticity in Equation 24. Indeed, all elasticity measures that we shall examine will have the same general structure, so essentially if you've seen one, you've seen them all. The only thing that changes is the independent variable of interest. For example, if the income elasticity of demand for some good has a value of 0.8, we would interpret that to mean that whenever income rises by one percent, the quantity demanded at each price would rise by 0.8 percent.

Although own-price elasticity of demand will almost always be negative because of the law of demand, *income* elasticity can be negative, positive, or zero. Positive income elasticity simply means that as income rises, quantity demanded also rises, as is characteristic of most consumption goods. We define a good with positive income elasticity as a **normal good**. It is perhaps unfortunate that economists often take perfectly good English words and give them different definitions. When an economist speaks of a normal good, he is saying nothing other than that the demand for that particular good rises when income increases and falls when income decreases. Hence, if we find that when income rises, people buy more meals at restaurants, then dining out is defined to be a normal good.

For some goods, there is an *inverse* relationship between quantity demanded and consumer income. That is, when people experience a rise in income, they buy absolutely less of some goods, and they buy more when their income falls. Hence, income elasticity of demand for those goods is negative. By definition, goods with *negative* income elasticity are called **inferior goods**. Again, the word inferior means nothing other than that the income elasticity of demand for that good is observed to be negative. It does not necessarily indicate anything at all about the quality of that good. Typical examples of inferior goods might be rice, potatoes, or less expensive cuts of meat. One study found that income elasticity of demand for beer is slightly negative, whereas income elasticity of demand for wine is significantly positive. An economist would therefore say that beer is inferior whereas wine is normal. Ultimately, whether

a good is called inferior or normal is simply a matter of empirical statistical analysis. And a good could be normal for one income group and inferior for another income group. (A BMW 3-series automobile might very well be normal for a moderate-income group but inferior for a high-income group of consumers. As their respective income levels rose, the moderate group might purchase more BMWs whereas the upper-income group might buy fewer 3-series as they traded up to a 5- or 7-series.) Clearly, for some goods and some ranges of income, consumer income might not have an impact on purchase decision at all. Hence for those goods, income elasticity of demand is zero.

Thinking back to our discussion of the demand curve, recall that we invoked the assumption of “holding all other things constant” when we plotted the relationship between price and quantity demanded. One of the variables we held constant was consumer income. If income were to change, obviously the whole curve would shift one way or the other. For normal goods, a rise in income would shift the entire demand curve upward and to the right, resulting in an *increase in demand*. If the good were inferior, however, a rise in income would result in a downward and leftward shift in the entire demand curve.

4.4 Cross-price Elasticity of Demand: Substitutes and Complements

It should be clear by now that any variable on the right-hand side of the demand function can serve as the basis for its own elasticity. Recall that the price of another good might very well have an impact on the demand for a good or service, so we should be able to define an elasticity with respect to the *other* price, as well. That elasticity is called the **cross-price elasticity of demand** and takes on the same structure as own-price elasticity and income elasticity of demand, as represented in Equation 26.

$$E_{P_y}^d = \frac{\% \Delta Q_x^d}{\% \Delta P_y} = \frac{\frac{\Delta Q_x^d}{Q_x^d}}{\frac{\Delta P_y}{P_y}} = \left(\frac{\Delta Q_x^d}{\Delta P_y} \right) \left(\frac{P_y}{Q_x^d} \right) \quad (26)$$

Note how similar in structure this equation is to own-price elasticity. The only difference is that the subscript on P is now y , indicating the price of some other good, Y , instead of the own-price, X . This cross-price elasticity of demand measures how sensitive the demand for good X is to changes in the price of some other good, Y , holding all other things constant. For some pairs of goods, X and Y , when the price of Y rises, more of good X is demanded. That is, the cross-price elasticity of demand is positive. Those goods are defined to be **substitutes**. Substitutes are defined empirically. If the cross-price elasticity of two goods is positive, they are substitutes, irrespective of whether someone would consider them “similar.”

This concept is intuitive if you think about two goods that are seen to be close substitutes, perhaps like two brands of beer. When the price of one of your favorite brands of beer rises, what would you do? You would probably buy less of that brand and more of a cheaper brand, so the cross-price elasticity of demand would be positive.

Alternatively, two goods whose cross-price elasticity of demand is negative are defined to be **complements**. Typically, these goods would tend to be consumed together as a pair, such as gasoline and automobiles or houses and furniture. When automobile prices fall, we might expect the quantity of autos demanded to rise, and thus we might expect to see a rise in the demand for gasoline. Ultimately, though, whether two goods are substitutes or complements is an empirical question answered solely by observation and statistical analysis. If, when the price of one good rises the demand for the other good also rises, they are substitutes. If the demand for that other good falls, they are complements. And the result might not immediately resonate with our

intuition. For example, grocery stores often put something like coffee on sale in the hope that customers will come in for coffee and end up doing their weekly shopping there as well. In that case, coffee and, say, cabbage could very well empirically turn out to be complements even though we do not normally think of consuming coffee and cabbage together as a pair (i.e., that the price of coffee has a relation to the sales of cabbage).

For substitute goods, an increase in the price of one good would shift the demand curve for the other good upward and to the right. For complements, however, the impact is in the other direction: When the price of one good rises, the quantity demanded of the other good shifts downward and to the left.

4.5 Calculating Demand Elasticities from Demand Functions

Although the concept of different elasticities of demand is helpful in sorting out the qualitative and directional effects among variables, the analyst will also benefit from having an empirically estimated demand function from which to calculate the magnitudes as well. There is no substitute for actual observation and statistical (regression) analysis to yield insights into the quantitative behavior of a market. (Empirical analysis, however, is outside the scope of this reading.) To see how an analyst would use such an equation, let us return to our hypothetical market demand function for gasoline in Equation 13 duplicated here:

$$Q_x^d = 8,400 - 400P_x + 60I - 10P_y \quad (27)$$

As we found when we calculated own-price elasticity of demand earlier, we need to identify “where to look” by choosing actual values for the independent variables, P_x , I , and P_y . We choose \$3 for P_x , \$50 (thousands) for I , and \$20 (thousands) for P_y . By inserting these values into the “estimated” demand function (Equation 27), we find that quantity demanded is 10,000 gallons of gasoline per week. We now have everything we need to calculate own-price, income, and cross-price elasticities of demand for our market. Those respective elasticities are expressed in Equations 28, 29, and 30. Each of those expressions has a term denoting the change in quantity divided by the change in each respective variable: $\Delta Q_x / \Delta P_x$, $\Delta Q_x / \Delta I$, and $\Delta Q_x / \Delta P_y$. In each case, those respective terms are given by the coefficients on the variables of interest. Once we recognize this fact, the rest is accomplished simply by inserting values into the elasticity formulas.

$$E_{P_x}^d = \left(\frac{\Delta Q_x^d}{\Delta P_x} \right) \left(\frac{P_x}{Q_x^d} \right) = [-400] \left[\frac{3}{10,000} \right] = -0.12 \quad (28)$$

$$E_I^d = \left(\frac{\Delta Q_x^d}{\Delta I} \right) \left(\frac{I}{Q_x^d} \right) = [60] \left[\frac{50}{10,000} \right] = 0.30 \quad (29)$$

$$E_{P_y}^d = \left(\frac{\Delta Q_x^d}{\Delta P_y} \right) \left(\frac{P_y}{Q_x^d} \right) = [-10] \left[\frac{20}{10,000} \right] = -0.02 \quad (30)$$

In our example, at a price of \$3, the own-price elasticity of demand is -0.12 , meaning that a 1 percent increase in the price of gasoline would bring about a decrease in quantity demanded of only 0.12 percent. Because the absolute value of the own-price elasticity is less than one, we characterize demand as being *inelastic* at that price, so an increase in price would result in an increase in total expenditure on gasoline by consumers in that market. Additionally, the income elasticity of demand is 0.30 , meaning that a 1 percent increase in income would bring about an increase of 0.30 percent in the quantity demanded of gasoline. Because that elasticity is positive (but small), we

would characterize gasoline as a *normal* good: An increase in income would cause consumers to buy more gasoline. Finally, the cross-price elasticity of demand between gasoline and automobiles is -0.02 , meaning that if the price of automobiles rose by 1 percent, the demand for gasoline would fall by 0.02 percent. We would therefore characterize gasoline and automobiles as *complements* because the cross-price elasticity is negative. The magnitude is, however, quite small, so we would conclude that the complementary relationship is quite weak.

EXAMPLE 13**Calculating Elasticities from a Given Demand Function**

An individual consumer's monthly demand for downloadable e-books is given by the equation $Q_{eb}^d = 2 - 0.4P_{eb} + 0.0005I + 0.15P_{hb}$, where Q_{eb}^d equals the number of e-books demanded each month, I equals the household monthly income, P_{eb} equals the price of e-books, and P_{hb} equals the price of hardbound books. Assume that the price of e-books is €10.68, household income is €2,300, and the price of hardbound books is €21.40.

- 1 Determine the value of own-price elasticity of demand for e-books.
- 2 Determine the income elasticity of demand for e-books.
- 3 Determine the cross-price elasticity of demand for e-books with respect to the price of hardbound books.

Solution to 1:

Recall that own-price elasticity of demand is given by $(\Delta Q_{eb}/\Delta P_{eb})(P_{eb}/Q_{eb})$, and notice from the demand function that $\Delta Q_{eb}/\Delta P_{eb} = -0.4$. Inserting the given variable values into the demand function yields $Q_{eb} = 2.088$. So at a price of €10.68, the own-price elasticity of demand equals $(-0.4)(10.68/2.088) = -2.046$, which is elastic because in absolute value the elasticity coefficient is greater than one.

Solution to 2:

Recall that income elasticity of demand is given by $(\Delta Q_{eb}/\Delta I)(I/Q_{eb})$. Notice from the demand function that $\Delta Q_{eb}/\Delta I = 0.0005$. Inserting in the values for I and Q_{eb} yields income elasticity of $(0.0005)(2,300/2.088) = 0.551$, which is positive, so e-books are a normal good.

Solution to 3:

Recall that cross-price elasticity of demand is given by $(\Delta Q_{eb}/\Delta P_{hb})(P_{hb}/Q_{eb})$, and notice from the demand function that $\Delta Q_{eb}/\Delta P_{hb} = 0.15$. Inserting in the values for P_{hb} and Q_{eb} yields a cross-price elasticity of demand for e-books of $(0.15)(21.40/2.088) = 1.537$, which is positive, implying that e-books and hard-bound books are substitutes.

SUMMARY

This reading has surveyed demand and supply analysis. Because markets (goods, factor, and capital) supply the foundation for today's global economy, an understanding of the demand and supply model is essential for any analyst who hopes to grasp the implications of economic developments on investment values. Among the points made are the following:

- The basic model of markets is the demand and supply model. The demand function represents buyers' behavior and can be depicted (in its inverse demand form) as a negatively sloped demand curve. The supply function represents sellers' behavior and can be depicted (in its inverse supply form) as a positively sloped supply curve. The interaction of buyers and sellers in a market results in equilibrium. Equilibrium exists when the highest price willingly paid by buyers is just equal to the lowest price willingly accepted by sellers.
- Goods markets are the interactions of consumers as buyers and firms as sellers of goods and services produced by firms and bought by households. Factor markets are the interactions of firms as buyers and households as sellers of land, labor, capital, and entrepreneurial risk-taking ability. Capital markets are used by firms to sell debt or equity to raise long-term capital to finance the production of goods and services.
- Demand and supply curves are drawn on the assumption that everything *except* the price of the good itself is held constant (an assumption known as *ceteris paribus* or "holding all other things constant"). When something other than price changes, the demand curve or the supply curve will shift relative to the other curve. This shift is referred to as a change in demand or supply, as opposed to quantity demanded or quantity supplied. A new equilibrium generally will be obtained at a different price and a different quantity than before. The market mechanism is the ability of prices to adjust to eliminate any excess demand or supply resulting from a shift in one or the other curve.
- If, at a given price, the quantity demanded exceeds the quantity supplied, there is excess demand and price will rise. If, at a given price, the quantity supplied exceeds the quantity demanded, there is excess supply and price will fall.
- Sometimes auctions are used to seek equilibrium prices. Common value auctions sell items that have the same value to all bidders, but bidders can only estimate that value before the auction is completed. Overly optimistic bidders overestimate the true value and end up paying a price greater than that value. This result is known as the winner's curse. Private value auctions sell items that (generally) have a unique subjective value for each bidder. Ascending price auctions use an auctioneer to call out ever increasing prices until the last, highest bidder ultimately pays his/her bid price and buys the item. Descending price, or Dutch, auctions begin at a very high price and then reduce that price until one bidder is willing to buy at that price. Second price sealed bid auctions are sometimes used to induce bidders to reveal their true reservation prices in private value auctions. Treasury notes and some other financial instruments are sold using a form of Dutch auction (called a single price auction) in which competitive and non-competitive bids are arrayed in descending price (increasing yield) order. The winning bidders all pay the same price, but marginal bidders might not be able to fill their entire order at the market clearing price.
- Markets that work freely can optimize society's welfare, as measured by consumer surplus and producer surplus. Consumer surplus is the difference between the total value to buyers and the total expenditure necessary to

purchase a given amount. Producer surplus is the difference between the total revenue received by sellers from selling a given amount and the total variable cost of production of that amount. When equilibrium price is reached, total surplus is maximized.

- Sometimes, government policies interfere with the free working of markets. Examples include price ceilings, price floors, and specific taxes. Whenever the imposition of such a policy alters the free market equilibrium quantity (the quantity that maximizes total surplus), there is a redistribution of surplus between buyers and sellers; but there is also a reduction of total surplus, called deadweight loss. Other influences can result in an imbalance between demand and supply. Search costs are impediments in the ability of willing buyers and willing sellers to meet in a transaction. Brokers can add value if they reduce search costs and match buyers and sellers. In general, anything that improves information about the willingness of buyers and sellers to engage will reduce search costs and add value.
- Economists use a quantitative measure of sensitivity called elasticity. In general, elasticity is the ratio of the percentage change in the dependent variable to the percentage change in the independent variable of interest. Important specific elasticities include own-price elasticity of demand, income elasticity of demand, and cross-price elasticity of demand.
- Based on algebraic sign and magnitude of the various elasticities, goods can be classified into groups. If own-price elasticity of demand is less than one in absolute value, demand is called “inelastic”; it is called “elastic” if own-price elasticity of demand is greater than one in absolute value. Goods with positive income elasticity of demand are called normal goods, and those with negative income elasticity of demand are called inferior goods. Two goods with negative cross-price elasticity of demand—a drop in the price of one good causes an increase in demand for the other good—are called complements. Goods with positive cross-price elasticity of demand—a drop in the price of one good causes a decrease in demand for the other—are called substitutes.
- The relationship among own-price elasticity of demand, changes in price, and changes in total expenditure is as follows: If demand is elastic, a reduction in price results in an increase in total expenditure; if demand is inelastic, a reduction in price results in a decrease in total expenditure; if demand is unitary elastic, a change in price leaves total expenditure unchanged.

PRACTICE PROBLEMS

- 1 Which of the following markets is *most* accurately characterized as a goods market? The market for:
 - A coats.
 - B sales clerks.
 - C cotton farmland.
- 2 The observation “as a price of a good falls, buyers buy more of it” is *best* known as:
 - A consumer surplus.
 - B the law of demand.
 - C the market mechanism.
- 3 Two-dimensional demand and supply curves are drawn under which of the following assumptions?
 - A Own price is held constant.
 - B All variables but quantity are held constant.
 - C All variables but own price and quantity are held constant.
- 4 The slope of a supply curve is *most* often:
 - A zero.
 - B positive.
 - C negative.
- 5 Assume the following equation

$$Q_x^s = -4 + \frac{1}{2}P_x - 2W$$

where Q_x^s is the quantity of good X supplied, P_x is the price of good X , and W is the wage rate paid to laborers. If the wage rate is 11, the vertical intercept on a graph depicting the supply curve is *closest* to:

- A -26.
 - B -4.
 - C 52.
- 6 Movement along the demand curve for good X occurs due to a change in:
 - A income.
 - B the price of good X .
 - C the price of a substitute for good X .

The following information relates to Questions 7–10

A producer's supply function is given by the equation

$$Q_s^s = -55 + 26P_s - 1.3W$$

where Q_s^s is the quantity of steel supplied by the market, P_s is the per unit price of steel, and W is the per unit price of labor.

- 7 If the price of labor rises, what happens to the steel producer's supply curve?
The supply curve:
- A shifts to the left.
 - B shifts to the right.
 - C remains unchanged.
- 8 If the unit price of labor is 10, which equation is *closest* to the expression for the inverse supply function?
- A $P_s = 2.6 + 0.04Q_s^s$.
 - B $Q_s^s = -42 + 26P_s$.
 - C $Q_s^s = -68 + 26P_s$.
- 9 If the unit price of labor is 10, the slope of the supply curve is *closest* to:
- A -1.30.
 - B 0.04.
 - C 26.00.
- 10 Assume the supply side of the market consists of exactly five identical sellers. If the unit price of labor is 20, which equation is *closest* to the expression for the market inverse supply function?
- A $P_s = 2.6 + 0.008Q_s^s$.
 - B $P_s = 3.1 + 0.008Q_s^s$.
 - C $P_s = 3.1 + 0.04Q_s^s$.

-
- 11 Which of the following statements about market equilibrium is *most* accurate?
- A The difference between quantity demanded and quantity supplied is zero.
 - B The demand curve is negatively sloped and the supply curve is positively sloped.
 - C For any given pair of market demand and supply curves, only one equilibrium point can exist.
- 12 Which of the following statements *best* characterizes the market mechanism for attaining equilibrium?
- A Excess supply causes prices to fall.
 - B Excess demand causes prices to fall.
 - C The demand and supply curves shift to reach equilibrium.
- 13 An auction in which the auctioneer starts at a high price and then lowers the price in increments until there is a willing buyer is *best* called a:

- A Dutch auction.
B Vickery auction.
C private-value auction.
- 14 Which statement is *most likely* to be true in a single-price US Treasury bill auction?
A Only some non-competitive bids would be filled.
B Bidders at the highest winning yield may only get a portion of their order filled.
C All bidders at a yield higher than the winning bid would get their entire order filled.
- 15 The winner's curse in common value auctions is *best* described as the winning bidder paying:
A more than the value of the asset.
B a price not equal to one's own bid.
C more than intended prior to bidding.
- 16 A wireless phone manufacturer introduced a next-generation phone that received a high level of positive publicity. Despite running several high-speed production assembly lines, the manufacturer is still falling short in meeting demand for the phone nine months after introduction. Which of the following statements is the *most* plausible explanation for the demand/supply imbalance?
A The phone price is low relative to the equilibrium price.
B Competitors introduced next-generation phones at a similar price.
C Consumer incomes grew faster than the manufacturer anticipated.
- 17 A per-unit tax on items sold that is paid by the seller will *most likely* result in the:
A supply curve shifting vertically upward.
B demand curve shifting vertically upward.
C demand curve shifting vertically downward.
- 18 Which of the following statements *most* accurately and completely describes a deadweight loss?
A A transfer of surplus from one party to another.
B A reduction in either the buyer's or seller's surplus.
C A reduction in total surplus resulting from market interference.
- 19 If an excise tax is paid by the buyer instead of the seller, which of the following statements is *most likely* to be true? The price (including tax):
A paid will be higher than if the seller had paid the tax.
B received will be lower than if the seller had paid the tax.
C received will be the same as if the seller had paid the tax.
- 20 A quota on an imported good below the market-clearing quantity will *most likely* lead to which of the following effects?
A The supply curve shifts upward.
B The demand curve shifts upward.
C Some of the buyer's surplus transfers to the seller.
- 21 Assume a market demand function is given by the equation

$$Q^d = 50 - 0.75P$$

where Q^d is the quantity demanded and P is the price. If P equals 10, the value of the consumer surplus is *closest* to:

- A 67.
 - B 1,205.
 - C 1,667.
- 22 Which of the following *best* describes producer surplus?
- A Revenue minus variable costs.
 - B Revenue minus variable plus fixed costs.
 - C The area above the supply curve and beneath the demand curve and to the left of the equilibrium point.
- 23 Assume a market supply function is given by the equation

$$Q_s = -7 + 0.6P$$

where Q_s is the quantity supplied and P is the price. If P equals 15, the value of the producer surplus is *closest* to:

- A 3.3.
- B 41.0.
- C 67.5.

The following information relates to Questions 24–26

The market demand function for four-year private universities is given by the equation

$$Q_{pr}^d = 84 - 3.1P_{pr} + 0.8I + 0.9P_{pu}$$

where Q_{pr}^d is the number of applicants to private universities per year in thousands, P_{pr} is the average price of private universities (in thousands of USD), I is the household monthly income (in thousands of USD), and P_{pu} is the average price of public (government-supported) universities (in thousands of USD). Assume that P_{pr} is equal to 38, I is equal to 100, and P_{pu} is equal to 18.

- 24 The price elasticity of demand for private universities is *closest* to:
- A -3.1.
 - B -1.9.
 - C 0.6.
- 25 The income elasticity of demand for private universities is *closest* to:
- A 0.5.
 - B 0.8.
 - C 1.3.
- 26 The cross-price elasticity of demand for private universities with respect to the price of public universities is *closest* to:
- A 0.3.

- B 3.1.
 - C 3.9.
-

27 If the cross-price elasticity between two goods is negative, the two goods are classified as:

- A normal.
- B substitutes.
- C complements.

SOLUTIONS

- 1 A is correct. Coats are finished goods, the result of the output of production.
- 2 B is correct.
- 3 C is correct. In order to draw demand and supply curves, own price and own quantity must be allowed to vary. However, all other variables are held constant to focus on the relation of own price with quantity.
- 4 B is correct. Producers generally will supply a greater quantity of a good at higher prices for the good.
- 5 C is correct. Because the supply curve is the graph of the inverse supply function, solve for the inverse supply function given the wage rate of 11:

$$\begin{aligned} Q_x^s &= -4 + \frac{1}{2}P_x - 2(11) \\ &= -26 + \frac{1}{2}P_x \\ Q_x^s + 26 &= \frac{1}{2}P_x \\ P_x &= 52 + 2Q_x^s \end{aligned}$$

The vertical intercept is 52.

- 6 B is correct. The demand curve shows quantity demanded as a function of own price only.
- 7 A is correct. The supply curve (which is the graph of the inverse supply function) shifts to the left. The producer is only willing to sell any given quantity at a higher price due to the increase in costs.
- 8 A is correct. The inverse supply function is *closest to* $P_s = 2.6 + 0.04Q_s^s$.

$$\begin{aligned} \text{Start with the supply equation: } Q_s^s &= -55 + 26P_s - 1.3W \\ \text{Insert } W = 10: \quad &= -55 + 26P_s - 1.3(10) \\ &= -68 + 26P_s \\ \text{Solve for } P_s: \quad P_s &= 2.6 + 0.04Q_s^s \text{ (the inverse supply function)} \end{aligned}$$

- 9 B is correct. The slope coefficient of Q_s^s in the inverse supply function, which gives the supply curve, is 0.04.
- 10 B is correct. Start with the equation $Q_s^s = -55 + 26P_s - 1.3W$. Insert the unit price of labor at 20 and, to aggregate for five suppliers, multiply the individual producer's supply function by 5:

$$\begin{aligned} Q_s^s &= 5(-55 + 26P_s - 1.3(20)) \\ Q_s^s &= -275 + 130P_s - 130 \\ Q_s^s &= -405 + 130P_s \end{aligned}$$

Invert the equation to get the market inverse supply function: $P_s = 3.1 + 0.008Q_s^s$.

- 11 A is correct. At market equilibrium the quantity demanded just equals quantity supplied, and thus, the difference between the two is zero.

- 12 A is correct. Excess supply at a given price implies that there is not enough demand at that price. So the price must fall until it reaches the point at which the demand and supply curves intersect.
- 13 A is correct. The basic Dutch auction is a descending-price auction.
- 14 B is correct. Non-competitive bids and bidders at lower yields will get their orders filled first. Securities may then not be available to fill demand entirely at the highest winning yield.
- 15 A is correct. The winning bidder in such auctions may be overly optimistic about the underlying value of the item won.
- 16 A is correct. The situation described is one of excess demand because, in order for markets to clear at the given level of quantity supplied, the company would need to raise prices.
- 17 A is correct. The lowest acceptable price to the supplier at any given quantity must now increase because part of the price is paid as a per-unit tax. Thus, the supply curve shifts upward.
- 18 C is correct. A deadweight loss is the surplus lost by both the producer and the consumer and not transferred to anyone.
- 19 C is correct. The trade price should be the same whether the tax is imposed on the buyer or on the seller.
- 20 C is correct. A quota will cause excess demand, raising the price of the good and moving it up and to the left along the demand curve. This should shift some of the buyer's surplus to the seller.
- 21 B is correct. We find consumer surplus as the area of the triangle formed by the y (price) axis, the inverse demand curve, and a line segment from the y axis to the inverse demand function at $P = 10$.

Put the price into the demand equation: $Q^d = 50 - 0.75(10)$
 $Q^d = 42.5$ (this is the base of the triangle)

Invert the demand function by solving for P : $-0.75P = Q^d - 50$

$$P = -1.33Q^d + 66.67$$

Note the price intercept is 66.67. The height of the triangle is $66.67 - 10 = 56.67$. The consumer surplus is the area of the triangle above the price of 10 and below the demand curve, with base equal to the quantity of 42.5: $\frac{1}{2} \text{Base} \times \text{Height} = (1/2)(42.5)(66.7 - 10) = 1,205$.

- 22 A is correct. Producer surplus is the difference between the total revenue that sellers receive from selling a given amount of a good and the total variable cost of producing that amount.
- 23 A is correct. With a linear supply curve, producer surplus is equal to the area of a triangle with base equal to the market clearing price minus the price intercept, height equal to the market clearing quantity, and bounded by the supply curve as the hypotenuse. Given a (market clearing) price of 15, quantity is 2:

$$Q_s = -7 + 0.6(15) = 2$$

Next find the inverse supply function:

$$P = (1/0.6)7 + (1/0.6)Q_s$$

$$P = 11.67 + 1.67Q_s$$

Note that the price intercept is 11.7 and the quantity intercept is -7.0. Thus, producer surplus is $\frac{1}{2} \text{Base} \times \text{Height} = (1/2)(2)(15 - 11.7) = 3.3$.

24 B is correct. From the demand function:

$$\Delta Q_{pr}^d / \Delta P_{pr} = -3.1 \text{ (the coefficient in front of own price)}$$

Solve for Q_{pr}^d :

$$\begin{aligned} Q_{pr}^d &= 84 - 3.1P_{pr} + 0.8I + 0.9P_{pu} \\ &= 84 - 3.1(38) + 0.8(100) + 0.9(18) \\ &= 62.4 \end{aligned}$$

$$\begin{aligned} \text{At } P_{pr} = 38, \text{ price elasticity of demand} &= (\Delta Q_{pr}^d / \Delta P_{pr})(P_{pr} / Q_{pr}^d) \\ &= (-3.1)(38/62.4) \\ &= -1.9 \end{aligned}$$

25 C is correct. From the demand function:

$$\Delta Q_{pr}^d / \Delta I = 0.8 \text{ (coefficient in front of the income variable)}$$

Solve for Q_{pr}^d :

$$\begin{aligned} Q_{pr}^d &= 84 - 3.1P_{pr} + 0.8I + 0.9P_{pu} \\ &= 84 - 3.1(38) + 0.8(100) + 0.9(18) \\ &= 62.4 \end{aligned}$$

$$\begin{aligned} \text{At } I = 100, \text{ the income elasticity of demand} &= (\Delta Q_{pr}^d / \Delta I)(I / Q_{pr}^d) \\ &= (0.8)(100/62.4) \\ &= 1.3 \end{aligned}$$

26 A is correct. From the demand function:

$$\Delta Q_{pr}^d / \Delta P_{pu} = 0.9 \text{ (the coefficient in front of } P_{pu})$$

Solve for Q_{pr}^d :

$$\begin{aligned} Q_{pr}^d &= 84 - 3.1P_{pr} + 0.8I + 0.9P_{pu} \\ &= 84 - 3.1(38) + 0.8(100) + 0.9(18) \\ &= 62.4 \end{aligned}$$

$$\begin{aligned} \text{At } P = 38, \text{ and } P_{pu} = 18, \text{ the cross-price elasticity of demand} &= (\Delta Q_{pr}^d / \Delta P_{pu})(P_{pu} / Q_{pr}^d) \\ &= (0.9)(18/62.4) \\ &= 0.3 \end{aligned}$$

27 C is correct. With complements, consumption goes up or down together. With a negative cross-price elasticity, as the price of one good goes up, the demand for both falls.



Glossary

- Arc elasticity** An elasticity based on two points, in contrast with (point) elasticity. With reference to price elasticity, the percentage change in quantity demanded divided by the percentage change in price between two points for price.
- Ascending price auction** An auction in which an auctioneer calls out prices for a single item and potential buyers bid directly against each other, with each subsequent bid being higher than the previous one.
- Behavioral equations** With respect to demand and supply, equations that model the behavior of buyers and sellers.
- Capital markets** Financial markets that trade securities of longer duration, such as bonds and equities.
- Change in quantity supplied** A movement along a given supply curve.
- Change in supply** A shift in the supply curve.
- Common value auction** An auction in which the item being auctioned has the same value to each auction participant, although participants may be uncertain as to what that value is.
- Complements** Said of goods which tend to be used together; technically, two goods whose cross-price elasticity of demand is negative.
- Consumer surplus** The difference between the value that a consumer places on units purchased and the amount of money that was required to pay for them.
- Consumption** The purchase of final goods and services by individuals.
- Cross-price elasticity of demand** The percent change in quantity demanded for a given small change in the price of another good; the responsiveness of the demand for Product A that is associated with the change in price of Product B.
- Deadweight loss** A net loss of total (consumer and producer) surplus.
- Demand** The willingness and ability of consumers to purchase a given amount of a good or service at a given price.
- Demand and supply analysis** The study of how buyers and sellers interact to determine transaction prices and quantities.
- Demand curve** Graph of the inverse demand function.
- Demand function** A relationship that expresses the quantity demanded of a good or service as a function of own-price and possibly other variables.
- Descending price auction** An auction in which the auctioneer begins at a high price, then lowers the called price in increments until there is a willing buyer for the item being auctioned.
- Dutch auction** An auction in which the auctioneer begins at a high price, then lowers the called price in increments until there is a willing buyer for the item being auctioned.
- Economics** The study of the production, distribution, and consumption of goods and services; the principles of the allocation of scarce resources among competing uses. Economics is divided into two broad areas of study: macroeconomics and microeconomics.
- Elastic** Said of a good or service when the magnitude of elasticity is greater than one.
- Elasticity** The percentage change in one variable for a percentage change in another variable; a measure of how sensitive one variable is to a change in the value of another variable.
- Endogenous variables** Variables whose equilibrium values are determined within the model being considered.
- Equilibrium condition** A condition necessary for the forces within a system to be in balance.
- Excess supply** A condition in which the quantity ready to be supplied is greater than the quantity demanded.
- Exogenous variables** Variables whose equilibrium values are determined outside of the model being considered.
- Externality** An effect of a market transaction that is borne by parties other than those who transacted.
- Factor markets** Markets for the purchase and sale of factors of production.
- First price sealed bid auction** An auction in which envelopes containing bids are opened simultaneously and the item is sold to the highest bidder.
- General equilibrium analysis** An analysis that provides for equilibria in multiple markets simultaneously.
- Goods markets** Markets for the output of production.
- Income elasticity of demand** A measure of the responsiveness of demand to changes in income, defined as the percentage change in quantity demanded divided by the percentage change in income.
- Inelastic** Insensitive to price changes.
- Inferior goods** A good whose consumption decreases as income increases.
- Intermediate goods and services** Goods and services purchased for use as inputs to produce other goods and services.
- Inverse demand function** A restatement of the demand function in which price is stated as a function of quantity.
- Labor markets** Markets for labor services.
- Law of demand** The principle that as the price of a good rises, buyers will choose to buy less of it, and as its price falls, they will buy more.
- Law of supply** The principle that a rise in price usually results in an increase in the quantity supplied.
- Macroeconomics** The branch of economics that deals with aggregate economic quantities, such as national output and national income.
- Marginal cost** The cost of producing an additional unit of a good.
- Marginal value** The added value from an additional unit of a good.
- Marginal value curve** A curve describing the highest price consumers are willing to pay for each additional unit of a good.
- Market equilibrium** The condition in which the quantity willingly offered for sale by sellers at a given price is just equal to the quantity willingly demanded by buyers at that same price.
- Market mechanism** The process by which price adjusts until there is neither excess supply nor excess demand.

Microeconomics The branch of economics that deals with markets and decision making of individual economic units, including consumers and businesses.

Negative externality A negative effect (e.g., pollution) of a market transaction that is borne by parties other than those who transacted; a spillover cost.

Normal good A good that is consumed in greater quantities as income increases.

Own-price The price of a good or service itself (as opposed to the price of something else).

Own-price elasticity of demand The percentage change in quantity demanded for a percentage change in own price, holding all other things constant.

Partial equilibrium analysis An equilibrium analysis focused on one market, taking the values of exogenous variables as given.

Perfectly elastic Said of a good or service that is infinitely sensitive to a change in the value of a specified variable (e.g., price).

Perfectly inelastic Said of a good or service that is completely insensitive to a change in the value of a specified variable (e.g., price).

Positive externality A positive effect (e.g., improved literacy) of a market transaction that is borne by parties other than those who transacted; a spillover benefit.

Price floor A minimum price for a good or service, typically imposed by government action and typically above the equilibrium price.

Private value auction An auction in which the value of the item being auctioned is unique to each bidder.

Producer surplus The difference between the total revenue sellers receive from selling a given amount of a good and the total variable cost of producing that amount.

Reservation prices The highest price a buyer is willing to pay for an item or the lowest price at which a seller is willing to sell it.

Saving In economics, income not spent.

Sealed bid auction An auction in which bids are elicited from potential buyers, but there is no ability to observe bids by other buyers until the auction has ended.

Search costs Costs incurred in searching; the costs of matching buyers with sellers.

Second price sealed bid An auction (also known as a Vickery auction) in which bids are submitted in sealed envelopes and opened simultaneously. The winning buyer is the one who submitted the highest bid, but the price paid is equal to the second highest bid.

Single price auction A Dutch auction variation, also involving a single price, is used in selling US Treasury securities.

Stable With reference to an equilibrium, one in which price, when disturbed away from the equilibrium, tends to converge back to it.

Substitutes Said of two goods or services such that if the price of one increases the demand for the other tends to increase, holding all other things equal (e.g., butter and margarine).

Supply The willingness of sellers to offer a given quantity of a good or service for a given price.

Supply curve The graph of the inverse supply function.

Supply function The quantity supplied as a function of price and possibly other variables.

Technology of production The “rules” that govern the transformation of inputs into finished goods and services.

Theory of the consumer The branch of microeconomics that deals with consumption—the demand for goods and services—by utility-maximizing individuals.

Theory of the firm The branch of microeconomics that deals with the supply of goods and services by profit-maximizing firms.

Total expenditure The total amount spent over a time period.

Total surplus The difference between total value to buyers and the total variable cost to sellers; made up of the sum of consumer surplus and producer surplus.

Unitary elastic An elasticity with a magnitude of 1.

Unit elastic An elasticity with a magnitude of 1.

Unstable With reference to an equilibrium, one in which price, when disturbed away from the equilibrium, tends not to return to it.

Variable costs Costs that fluctuate with the level of production and sales.

Winner's curse The tendency for the winner in certain competitive bidding situations to overpay, whether because of overestimation of intrinsic value, emotion, or information asymmetries.

READING

14

Demand and Supply Analysis: Consumer Demand

by Richard V. Eastin, PhD, and Gary L. Arbogast, CFA

Richard V. Eastin, PhD, is at the University of Southern California (USA). Gary L. Arbogast, CFA (USA).

LEARNING OUTCOMES

Mastery	<i>The candidate should be able to:</i>
<input type="checkbox"/>	a. describe consumer choice theory and utility theory;
<input type="checkbox"/>	b. describe the use of indifference curves, opportunity sets, and budget constraints in decision making;
<input type="checkbox"/>	c. calculate and interpret a budget constraint;
<input type="checkbox"/>	d. determine a consumer's equilibrium bundle of goods based on utility analysis;
<input type="checkbox"/>	e. compare substitution and income effects;
<input type="checkbox"/>	f. distinguish between normal goods and inferior goods and explain Giffen goods and Veblen goods in this context.

INTRODUCTION

1

By now it should be clear that economists are model builders. In the previous reading, we examined one of their most fundamental models, the model of demand and supply. And as we have seen, models begin with simplifying assumptions and then find the implications that can then be compared to real-world observations as a test of the model's usefulness. In the model of demand and supply, we *assumed* the existence of a demand curve and a supply curve, as well as their respective negative and positive slopes. That simple model yielded some very powerful implications about how markets work, but we can delve even more deeply to explore the underpinnings of demand and supply. In this reading, we examine the theory of the consumer as a way of understanding where consumer demand curves originate. In a subsequent reading, the origins of the supply curve are sought in presenting the theory of the firm.

This reading is organized as follows: Section 2 describes consumer choice theory in more detail. Section 3 introduces utility theory, a building block of consumer choice theory that provides a quantitative model for a consumer's preferences and tastes. Section 4 surveys budget constraints and opportunity sets. Section 5 covers the determination of the consumer's bundle of goods and how that may change in response to changes in income and prices. Section 6 examines substitution and income effects for different types of goods. A summary and practice problems conclude the reading.

2

CONSUMER THEORY: FROM PREFERENCES TO DEMAND FUNCTIONS

The introduction to demand and supply analysis in the previous reading basically assumed that the demand function exists, and focused on understanding its various characteristics and manifestations. In this reading, we address the foundations of demand and supply analysis and seek to understand the sources of consumer demand through the theory of the consumer, also known as consumer choice theory. **Consumer choice theory** can be defined as the branch of microeconomics that relates consumer demand curves to consumer preferences. Consumer choice theory begins with a fundamental model of how consumer preferences and tastes might be represented. It explores consumers' willingness to trade off between two goods (or two baskets of goods), both of which the consumer finds beneficial. Consumer choice theory then recognizes that to consume a set of goods and services, consumers must purchase them at given market prices and with a limited income. In effect, consumer choice theory first models what the consumer would like to consume, and then it examines what the consumer can consume with limited income. Finally, by superimposing what the consumer would *like* to do onto what the consumer *can* do, we arrive at a model of what the consumer *would* do under various circumstances. Then by changing prices and income, the model develops consumer demand as a logical extension of consumer choice theory.

Although consumer choice theory attempts to model consumers' preferences or tastes, it does not have much to say about *why* consumers have the tastes and preferences they have. It still makes assumptions, but does so at a more fundamental level. Instead of assuming the existence of a demand curve, it derives a demand curve as an implication of assumptions about preferences. Note that economists are not attempting to predict the behavior of any single consumer in any given circumstance. Instead, they are attempting to build a consistent model of aggregate market behavior in the form of a market demand curve.

Once we model the consumer's preferences, we then recognize that consumption is governed not only by preferences but also by the consumer's **budget constraint** (the ability to purchase various combinations of goods and services, given his or her income). Putting preference theory together with the budget constraint gives us the demand curve we are seeking. In the following sections, we explore these topics in turn.

UTILITY THEORY: MODELING PREFERENCES AND TASTES

3

At the foundation of consumer behavior theory is the assumption that the consumer knows his or her own tastes and preferences and tends to take rational actions that result in a more preferred consumption “bundle” over a less preferred bundle. To build a consistent model of consumer choice, we need to begin with a few assumptions about preferences.

3.1 Axioms of the Theory of Consumer Choice

First, let us be clear about the consumption opportunities over which the consumer is assumed to have preferences. We define a **consumption bundle** or **consumption basket** as a specific combination of the goods and services that the consumer would like to consume. We could almost literally conceive of a basket containing a given amount of, say, shoes, pizza, medical care, theater tickets, piano lessons, and all the other things that a consumer might enjoy consuming. Each of those goods and services can be represented in a given basket by a non-negative quantity, respectively, of all the possible goods and services. Any given basket could have zero of one or more of those goods. A distinctly different consumption bundle would contain all of the same goods but in different quantities, again allowing for the possibility of a zero quantity of one or more of the goods. For example, bundle *A* might have the same amount of all but one of the goods and services as bundle *B* but a different amount of that one. Bundles *A* and *B* would be considered two distinct bundles.

Given this understanding of consumption bundles, the first assumption we make about a given consumer’s preferences is simply that she is able to make a comparison between any two possible bundles. That is, given bundles *A* and *B*, she must be able to say either that she prefers *A* to *B*, or she prefers *B* to *A*, or she is *indifferent* between the two. This is the assumption of **complete preferences** (also known as the axiom of completeness), and although it does not appear to be a particularly strong assumption, it is not trivial either. It rules out the possibility that she could just say, “I recognize that the two bundles are different, but in fact they are *so* different that I simply cannot compare them at all.” A loving father might very well say that about his two children. In effect, the father neither prefers one to the other nor is, in any meaningful sense, indifferent between the two. The assumption of complete preferences cannot accommodate such a response.

Second, we assume that when comparing any three distinct bundles, *A*, *B*, and *C*, if *A* is preferred to *B*, and simultaneously *B* is preferred to *C*, then it must be true that *A* is preferred to *C*. This assumption is referred to as the assumption of **transitive preferences**, and it is assumed to hold for indifference as well as for strict preference. This is a somewhat stronger assumption because it is essentially an assumption of rationality. We would say that if a consumer prefers a skiing holiday to a diving holiday and a diving holiday to a backpacking holiday and at the same time prefers a backpacking holiday to a skiing holiday, then he is acting irrationally. Transitivity rules out this kind of inconsistency. If you have studied psychology, however, you will no doubt have seen experiments that show subjects violating this assumption, especially in cases of many complex options being offered to them.

When we state these axioms, we are not saying that we believe them actually to be true in every instance, but we assume them for the sake of building a model. A model is a simplification of the real world phenomena we are trying to understand. Necessarily, axioms must be at some level inaccurate and incomplete representations

of the phenomena we are trying to model. If that were not the case, the “model” would not be a simplification; it would be a reflection of the complex system we are attempting to model and thus would not help our understanding very much.

Finally, we usually assume that in at least one of the goods, the consumer could never have so much that she would refuse any more, even if it were free. This assumption is sometimes referred to as the “more is better” assumption or the assumption of **non-satiation**. Clearly, for some things, more *is* worse, such as air pollution or trash. In those cases, the *good* is then the *removal* of that *bad*, so we can usually reframe our model to accommodate the non-satiation assumption. In particular, when we later discuss the concept of risk for an investor, we will recognize that for many, more risk is worse than less risk, all else being equal. In that analysis, we shall model the willingness of the investor to trade off between increased investment returns and increased certainty, which is the absence of risk.

EXAMPLE 1

Axioms Concerning Preferences

Helen Smith enjoys, among other things, eating sausages. She also enjoys reading Marcel Proust. Smith is confronted with two baskets: Basket A, which contains several other goods and a package of sausages, and B, which contains identical quantities of the other goods as Basket A, but instead of the sausages, it contains a book by Proust. When asked which basket she prefers, she replies, “I like them both, but sausages and a book by Proust are *so* different that I simply cannot compare the two baskets.” Determine whether Smith is obeying all the axioms of preference theory.

Solution:

Smith is violating the assumption of complete preferences. This assumption states that a consumer must be able to compare any two baskets of goods, either preferring one to the other or being indifferent between the two. If she complies with this assumption, she must be able to compare these two baskets of goods.

3.2 Representing the Preference of a Consumer: The Utility Function

Armed with the assumptions of completeness, transitivity, and non-satiation, we ask whether there might be a way for a given consumer to represent his own preferences in a consistent manner. Let us consider presenting him with all possible bundles of all the possible goods and services he could consider. Now suppose we give him paper and pencil and ask him to assign a number to each of the bundles. (The assumption of completeness ensures that he, in fact, could do that.) All he must do is write a number on a paper and lay it on each of the bundles. The only restrictions are these: Comparing any two bundles, if he prefers one to the other, he must assign a higher number to the bundle he prefers. And if he is indifferent between them, he must assign the same number to both. Other than that, he is free to begin with any number he wants for the first bundle he considers. In this way, he is simply ordering the bundles according to his preferences over them.

Of course, each of these possible bundles has a specific quantity of each of the goods and services. So, we have two sets of numbers. One set consists of the pieces of paper he has laid on the bundles. The other is the set of numerical quantities of the goods that are contained in each of the respective bundles. Under “reasonable assumptions” (the definition of which is not necessary for us to delve into at this level), it is possible

to come up with a rule that translates the quantities of goods in each basket into the number that our consumer has assigned to each basket. That “assignment rule” is called the **utility function** of that particular consumer. The single task of that utility function is to translate each basket of goods and services into a number that rank orders the baskets according to our particular consumer’s preferences. The number itself is referred to as the utility of that basket and is measured in **utils**, which are just quantities of happiness, or well-being, or whatever comes to mind such that more of it is better than less of it.

In general, we can represent the utility function as

$$U = f(Q_{x_1}, Q_{x_2}, \dots, Q_{x_n}) \quad (1)$$

where the Q s are the quantities of each of the respective goods and services in the bundles. In the case of two goods—say, ounces of wine (W) and slices of bread (B)—a utility function might be simply

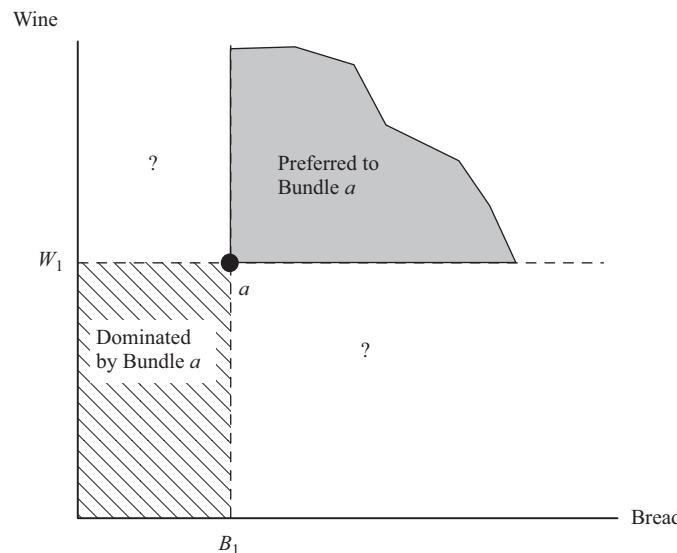
$$U = f(W, B) = WB \quad (2)$$

or the product of the number of ounces of wine and the number of slices of bread. The utility of a bundle containing 4 ounces of wine along with 2 slices of bread would equal 8 utils, and it would rank lower than a bundle containing 3 ounces of wine along with 3 slices of bread, which would yield 9 utils.

The important point to note is that the utility function is just a ranking of bundles of goods. If someone were to replace all those pieces of paper with new numbers that maintained the same ranking, then the new set of numbers would be just as useful a utility function as the first in describing our consumer’s preferences. This characteristic of utility functions is called an *ordinal*, as contrasted to a *cardinal*, ranking. Ordinal rankings are weaker measures than cardinal rankings because they do not allow the calculation and ranking of the *differences* between bundles.

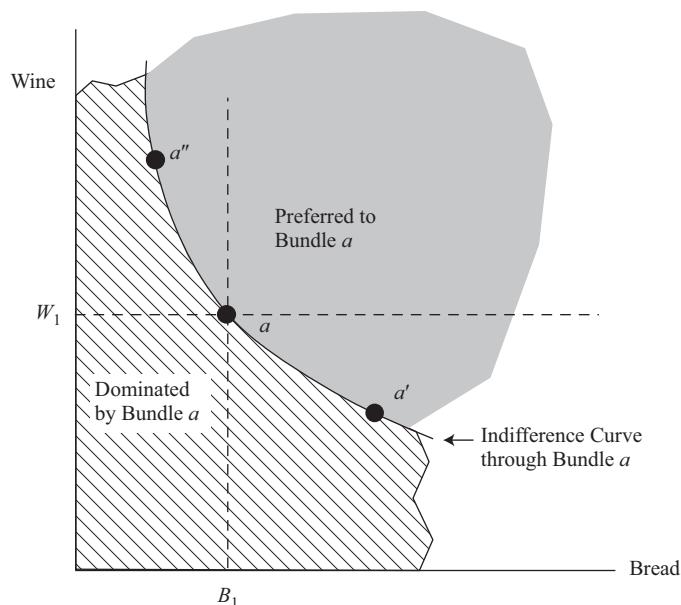
3.3 Indifference Curves: The Graphical Portrayal of the Utility Function

It will be convenient for us to represent our consumer’s preferences graphically, not just mathematically. To that end, we introduce the concept of an **indifference curve**, which represents all the combinations of two goods such that the consumer is entirely indifferent among them. This is how we construct such a curve: Consider bundles that contain only two goods so that we can use a two-dimensional graph to represent them—as in Exhibit 1, where a particular bundle containing W_a ounces of wine along with B_a slices of bread is represented as a single point, a . The assumption of non-satiation (more is always better) ensures that all bundles lying directly above, directly to the right of, or both above and to the right (more wine and more bread) of point a must be preferred to bundle a . That set of bundles is called the “preferred-to-bundle- a ” set. Correspondingly, all the bundles that lie directly below, to the left of, and both below and to the left of bundle a must yield less utility and therefore would be called the “dominated-by-bundle- a ” set.

Exhibit 1 Showing Preferences Graphically


Note: A given bundle of two goods is represented as a single point, a , in the two-dimensional graph. Non-satiation allows us to rank-order many, but not all, other bundles, relative to a , leaving some questions unanswered.

To determine our consumer's preferences, suppose we present a choice between bundle a and some bundle a' , which contains more bread but less wine than a . Non-satiation is not helpful to us in this case, so we need to ask the consumer which he prefers. If he strictly prefers a' , then we would remove a little bread and ask again. If he strictly prefers a , then we would add a little bread, and so on. Finally, after a series of adjustments, we could find just the right combination of bread and wine such that the new bundle a' would be equally satisfying to our consumer as bundle a . That is to say, our consumer would be indifferent between consuming bundle a or bundle a' . We would then choose a bundle, say a'' , that contains more wine and less bread than bundle a , and we would again adjust the goods such that the consumer is once again indifferent between bundle a and bundle a'' . By continuing to choose bundles and make adjustments, it would be possible to identify all possible bundles such that the consumer is just indifferent among each of them and bundle a . Such a set of points is represented in Exhibit 2, where the indifference curve through point a represents that set of bundles. Notice that the "preferred-to-bundle- a " set has expanded to include all bundles that lie in the region above and to the right of the indifference curve. Correspondingly, the "dominated-by-bundle- a " set has expanded to include all bundles that lie in the region below and to the left of the indifference curve.

Exhibit 2 An Indifference Curve

Note: An indifference curve shows all combinations of two goods such that the consumer is indifferent between them.

The indifference curve represents our consumer's unique preferences over the two goods wine and bread. Its negative slope simply represents that both wine and bread are seen as "good" to this consumer; in order to maintain indifference, a decrease in the quantity of wine must be compensated for by an increase in the quantity of bread. Its curvature tells us something about the strength of his willingness to trade off one good for the other. The indifference curve in Exhibit 2 is characteristically drawn to be *convex* when viewed from the origin. This indicates that the willingness to give up wine to obtain a little more bread diminishes the more bread and the less wine the bundle contains.

We capture this willingness to give up one good to obtain a little more of the other in the phrase **marginal rate of substitution** of bread for wine, MRS_{BW} . The MRS_{BW} is the rate at which the consumer is willing to give up wine to obtain a small increment of bread, holding utility constant (i.e., movement along an indifference curve). Notice that the convexity implies that at a bundle like a'' , which contains rather a lot of wine and not much bread, the consumer would be willing to give up a considerable amount of wine in exchange for just a little more bread. (The slope of the indifference curve is quite steep at that point.) However, at a point like a' , which contains considerably more bread but less wine than a'' , the consumer is not ready to sacrifice nearly as much wine to obtain a little more bread. This suggests that the value being placed on bread, in terms of the amount of wine the consumer is willing to give up for bread, diminishes the more bread and less wine he has. It follows that the MRS_{BW} is the negative of the slope of the tangent to the indifference curve at any given bundle. If, at some point, the slope of the indifference curve had value -2.5 , it means that, starting at that particular bundle, our consumer would be willing to sacrifice wine to obtain bread at the rate of 2.5 ounces of wine per slice of bread. Because of the convexity assumption—that MRS_{BW} must diminish as he moves toward more bread and less wine—the MRS_{BW} is continuously changing as he moves along his indifference curve.

EXAMPLE 2**Understanding the Marginal Rate of Substitution**

Tom Warren currently has 50 blueberries and 20 peanuts. His marginal rate of substitution of peanuts for blueberries, MRS_{pb} , equals 4, and his indifference curves are strictly convex.

- 1 Determine whether Warren would be willing to trade at the rate of 3 of his blueberries in exchange for 1 more peanut.
- 2 Suppose that Warren is indifferent between his current bundle and one containing 40 blueberries and 25 peanuts. Describe Warren's MRS_{pb} evaluated at the new bundle.

Solution to 1:

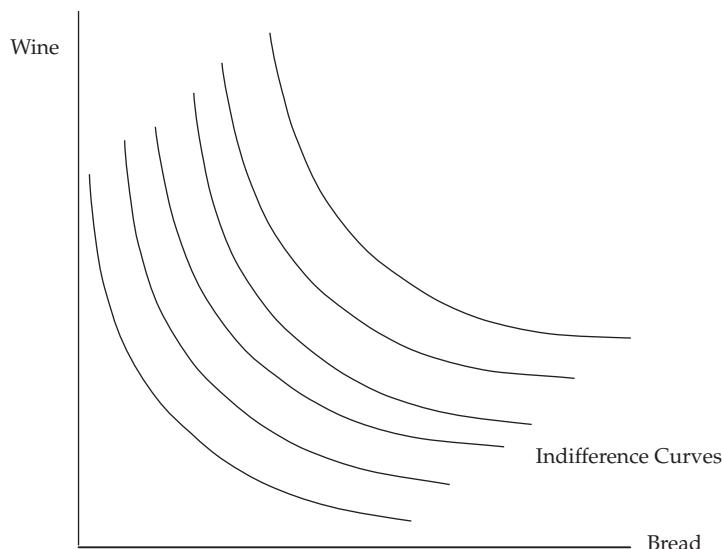
$MRS_{pb} = 4$ means that Warren would be willing to give up 4 blueberries for 1 peanut, at that point. He clearly would be willing to give up blueberries at a rate less than that, namely, 3-to-1.

Solution to 2:

The new bundle has more peanuts and fewer blueberries than the original one, and Warren is indifferent between the two, meaning that both bundles lie on the same indifference curve, where blueberries are plotted on the vertical axis and peanuts on the horizontal axis. Because his indifference curves are strictly convex and the new bundle lies below and to the right of his old bundle, his MRS_{pb} must be less than 4. That is to say, his indifference curve at the new point must be less steep than at the original bundle.

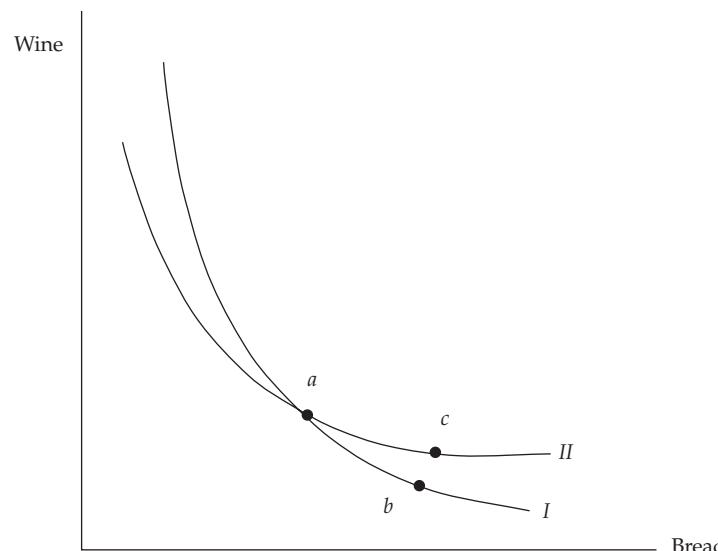
3.4 Indifference Curve Maps

There was nothing special about our initial choice of bundle α as a starting point for the indifference curve. We could have begun with a bundle containing more of both goods. In that case, we could have gone through the same process of trial and error, and we would have ended up with another indifference curve, this one passing through the new point and lying above and to the right of the first one. Indeed, we could construct any number of indifference curves in the same manner simply by starting at a different initial bundle. The result is an entire family of indifference curves, called an **indifference curve map**, and it represents our consumer's entire utility function. The word *map* is appropriate because the entire set of indifference curves comprises a contour map of this consumer's utility function. Each contour, or indifference curve, is a set of points in which each point shares a common level of utility with the others. Moving upward and to the right from one indifference curve to the next represents an increase in utility, and moving down and to the left represents a decrease. The map could look like that in Exhibit 3.

Exhibit 3 An Indifference Curve Map

Note: The indifference curve map represents the consumer's utility function. Any curve above and to the right represents a higher level of utility.

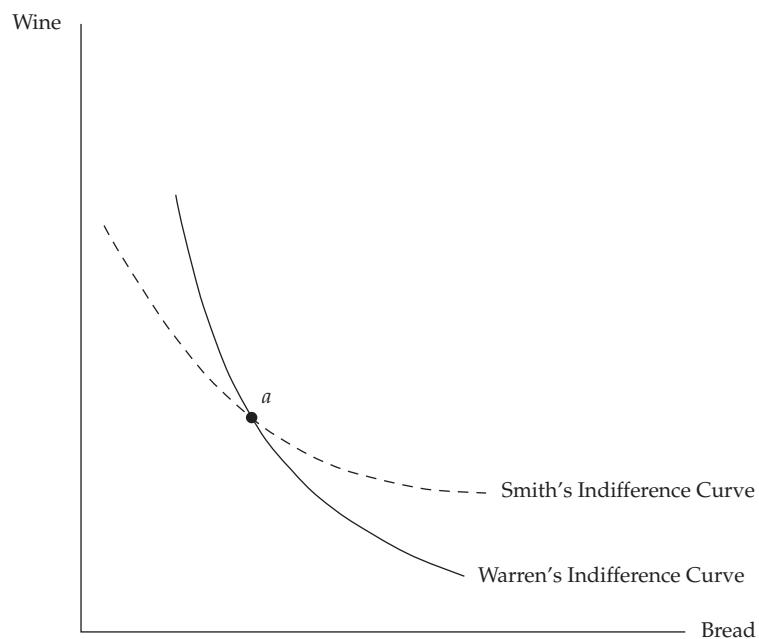
Because of the completeness assumption, there will be one indifference curve passing through every point in the set. Because of the transitivity assumption, no two indifference curves for a given consumer can ever cross. Exhibit 4 shows why. If bundle a and bundle b lie on the same indifference curve, the consumer must be indifferent between the two. If a and c lie on the same indifference curve, she must be indifferent between these two bundles as well. But because bundle c contains more of both wine and bread than bundle b , she must prefer c to b , which violates transitivity of preferences. So we see that indifference curves will generally be strictly convex and negatively sloped, and they cannot cross. These are the only restrictions we place on indifference curve maps.

Exhibit 4 Why One Person's Indifference Curves Cannot Cross

Note: Two indifference curves for a given individual cannot cross because the transitivity assumption would be violated.

3.5 Gains from Voluntary Exchange: Creating Wealth through Trade

There is no requirement that all consumers have the same preferences. Take the case of Helen Smith and Tom Warren. The indifference curves for Smith will likely be different from Warren's. And although *for any given individual* two indifference curves cannot cross, there is no reason why two indifference curves for two different consumers cannot intersect. Consider Exhibit 5, in which we observe an indifference curve for Smith and one for Warren. Suppose they are initially endowed with identical bundles, represented by *a*. They each have exactly identical quantities of bread and wine. Note, however, that because their indifference curves intersect at that point, their slopes are different. Warren's indifference curve is steeper at point *a* than is Smith's. This means that Warren's MRS_{BW} is greater than Smith's MRS_{BW} . That is to say, Warren is willing, at that point, to give up more wine for an additional slice of bread than Smith is. That also means that Smith is willing to give up more bread for an additional ounce of wine than Warren is. Therefore, we observe that Warren has a relatively stronger preference for bread compared to Smith, and Smith has a relatively stronger preference for wine than Warren.

Exhibit 5 Two Consumers with Different Preferences

Note: When two consumers have different preferences, they will have different marginal rates of substitution when evaluated at identical bundles. Here, Warren has a relatively strong preference for bread because he is willing to give up more wine for another slice of bread than is Smith.

Suppose that the slope of Warren's indifference curve at point a is equal to -2 , and the slope of Smith's indifference curve at point a is equal to $-\frac{1}{2}$. Warren is willing to give up 2 ounces of wine for 1 slice of bread, and Smith is willing to give up only $\frac{1}{2}$ ounce of wine for 1 slice of bread. But that means she would be willing to give up 2 slices of bread for 1 ounce of wine. What would happen if Warren and Smith are allowed to exchange bread for wine? Suppose they are allowed to exchange at the ratio of one ounce of wine for one slice of bread. Would they both agree to an exchange at that ratio? Yes. Warren would be willing to give up two ounces of wine for a slice of bread, so he would certainly be willing to give up only one ounce of wine for one slice of bread. Correspondingly, Smith would be willing to give up two slices of bread for one ounce of wine, so she would certainly be willing to give up only one slice of bread for one ounce of wine. If they actually made such a trade at the one-to-one ratio, then Smith would end up with more wine and less bread than she started with, and Warren would end up with more bread and less wine than he started with.

We could say that Warren is better off by the value to him of one ounce of wine because he was *willing* to give up two ounces but only *had to* give up one ounce for his slice of bread. What about Smith? She is better off by the value to her of one slice of bread because she was *willing* to give up two slices of bread for her one additional ounce of wine but only *had to* give up one slice. Both Smith and Warren are better off after they trade. There is no more bread or wine than when they began, but there is greater *wealth* because both are better off than before they traded with each other. Both Smith and Warren ended on higher indifference curves than when they began.

As Smith gives up slices of bread for more ounces of wine, her MRS_{BW} increases; her indifference curve becomes steeper. Simultaneously, as Warren gives up ounces of wine for more slices of bread, his MRS_{BW} decreases; his indifference curve becomes less steep. Eventually, if they continue to trade, their MRS s will reach equality and there will be no further gains to be achieved from additional exchange. Initially, it

was the differences in their willingness to trade one good for the other that made trading beneficial to both. But if they trade to a pair of bundles at which their MRSs are equal, then trading will cease.

EXAMPLE 3

Understanding Voluntary Exchange

Helen Smith and Tom Warren have identical baskets containing books (B) and compact discs (D). Smith's MRS_{BD} equals 0.8 (i.e., she is willing to give up 0.8 disc for 1 book), and Warren's MRS_{BD} equals 1.25.

- 1 Determine whether Warren would accept the trade of 1 of Smith's discs in exchange for 1 of his books.
- 2 State and justify whether Smith or Warren has a relatively stronger preference for books.
- 3 Determine whether Smith or Warren would end up with more discs than he/she had to begin with, assuming they were allowed to exchange at the rate of 1 book for 1 disc. Justify your answer.

Solution to 1:

Warren's MRS_{BD} equals 1.25, meaning that he is willing to give up 1.25 discs for 1 more book. Another way to say this is that Warren requires at least 1.25 discs to compensate him for giving up 1 book. Because Smith only offers one disc, Warren will not accept the offer. (Of course, Smith would not voluntarily give up one disc for one of Warren's books. Her MRS_{BD} is only 0.8, meaning that she would be willing to give up, at most, 0.8 disc for a book; so she would not have offered one disc for a book anyway.)

Solution to 2:

Because Warren is willing to give up 1.25 discs for a book and Smith is willing to give up only 0.8 disc for a book, Warren has a relatively stronger preference for books.

Solution to 3:

Smith would have more discs than she originally had. Because Smith has a relatively stronger preference for discs and Warren has a relatively stronger preference for books, Smith would trade books for discs and so would end up with more discs.

4

THE OPPORTUNITY SET: CONSUMPTION, PRODUCTION, AND INVESTMENT CHOICE

Above, we have examined the trade-offs that economic actors (e.g., consumers, companies, investors) are *willing* to make. In this section, we recognize that circumstances almost always impose constraints on the trade-offs that these actors *are able* to make. In other words, we need to explore how to model the constraints on behavior that are imposed by the fact that we live in a world of scarcity: There is simply not enough of everything to satisfy the needs and desires of everyone at a given time. Consumers must generally purchase goods and services with their limited incomes and at given market prices. Companies, too, must divide their limited input resources in order to

produce different products. Investors are not able to choose *both* high returns *and* low risk simultaneously. Choices must be made, and here we examine how to represent the set of choices from which to choose.

4.1 The Budget Constraint

Previously, we examined what would happen if Warren and Smith were each given an endowment of bread and wine and were allowed to exchange at some pre-determined ratio. Although that circumstance is possible, a more realistic situation would be if Warren or Smith had a given income with which to purchase bread and wine at fixed market prices. Let Warren's income be given by I , the price he must pay for a slice of bread be P_B , and the price he must pay for an ounce of wine be P_W . Warren has freedom to spend his income any way he chooses, so long as the expenditure on bread plus the expenditure on wine does not exceed his income per time period. We can represent this **income constraint** (or **budget constraint**) with the following expression:

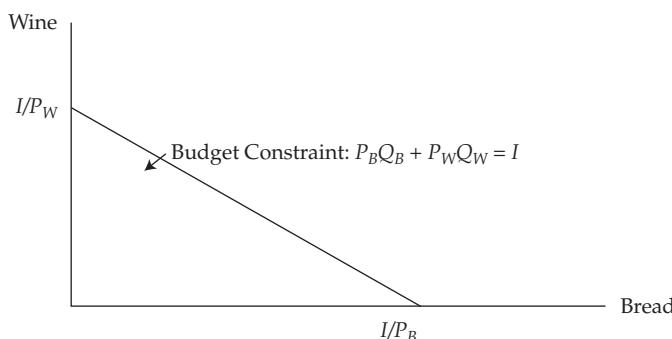
$$P_B Q_B + P_W Q_W \leq I \quad (3)$$

This expression simply constrains Warren to spend, in total, no more than his income. At this stage of our analysis, we are assuming a one-period model. In effect, then, Warren has no reason *not* to spend all of his income. The weak inequality becomes a strict equality, as shown in Equation 4, because there would be no reason for Warren to save any of his income if there is "no tomorrow."

$$P_B Q_B + P_W Q_W = I \quad (4)$$

From this equation, we see that if Warren were to spend all of his income only on bread, he could buy I/P_B slices of bread. Or if he were to confine his expenditure to wine alone, he could buy I/P_W ounces of wine. Alternatively, he could spread his income across bread and wine expenditures any way he chooses. Graphically, then, his budget constraint would appear as in Exhibit 6:

Exhibit 6 The Budget Constraint



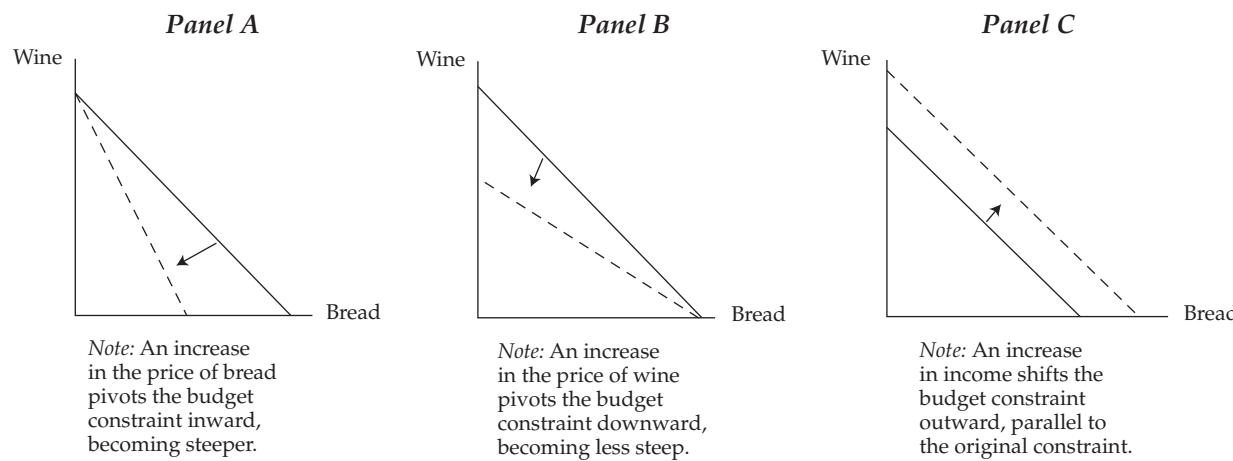
Note: The budget constraint shows all the combinations of bread and wine that the consumer could purchase with a fixed amount of income, I , paying prices P_B and P_W , respectively.

A simple algebraic manipulation of Equation 4 yields the budget constraint in the form of an intercept and slope:

$$Q_W = \frac{I}{P_W} - \frac{P_B}{P_W} Q_B \quad (5)$$

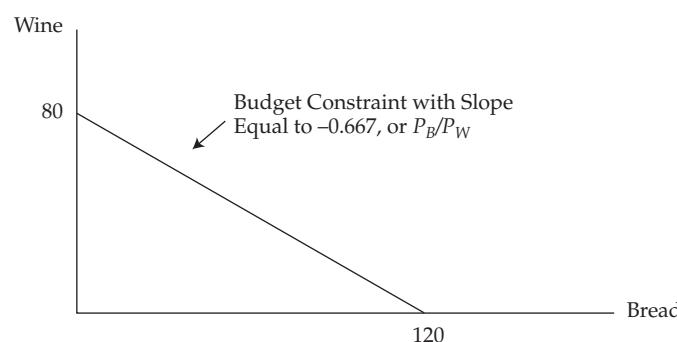
Notice that the slope of the budget constraint is equal to $-P_B/P_W$, and it shows the amount of wine that Warren would have to give up if he were to purchase another slice of bread. If the price of bread were to rise, the budget constraint would become steeper, pivoting through the vertical intercept. Alternatively, if the price of wine were to rise, the budget constraint would become less steep, pivoting downward through the horizontal intercept. If income were to rise, the entire budget constraint would shift outward, parallel to the original constraint, as shown in Exhibit 7:

Exhibit 7 Changing Prices and Income



As a specific example of a budget constraint, suppose Smith has \$60 to spend on bread and wine per month, the price of a slice of bread is \$0.50, and the price of an ounce of wine is \$0.75. If she spent all of her income on bread, she could buy 120 slices of bread. Or she could buy up to 80 ounces of wine if she chose to buy no bread. Obviously, she can spend half her income on each good, in which case she could buy 60 slices of bread and 40 ounces of wine. The entire set of bundles that Smith could buy with her \$60 budget is shown in Exhibit 8:

Exhibit 8 A Specific Example of a Budget Constraint



Note: This exhibit shows Smith's budget constraint if she has an income of \$60 and must pay \$0.50 per slice of bread and \$0.75 per ounce of wine.

EXAMPLE 4**The Budget Constraint**

Nigel's Pub has a total budget of £128 per week to spend on cod and lamb. The price of cod is £16 per kilogram, and the price of lamb is £10 per kilogram.

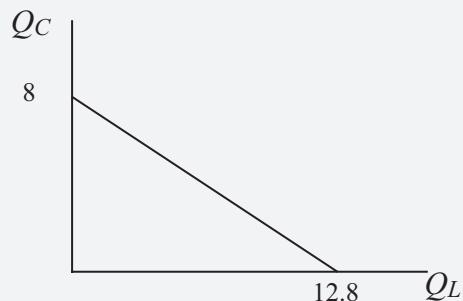
- 1 Calculate Nigel's budget constraint.
- 2 Construct a diagram of Nigel's budget constraint.
- 3 Determine the slope of Nigel's budget constraint.

Solution to 1:

The budget constraint is simply that the sum of the expenditure on cod plus the expenditure on lamb be equal to his budget: $128 = 16 Q_C + 10 Q_L$. Rearranging, it can also be written in intercept slope form: $Q_C = 128/P_C - (P_L/P_C) Q_L = 8 - 0.625 Q_L$.

Solution to 2:

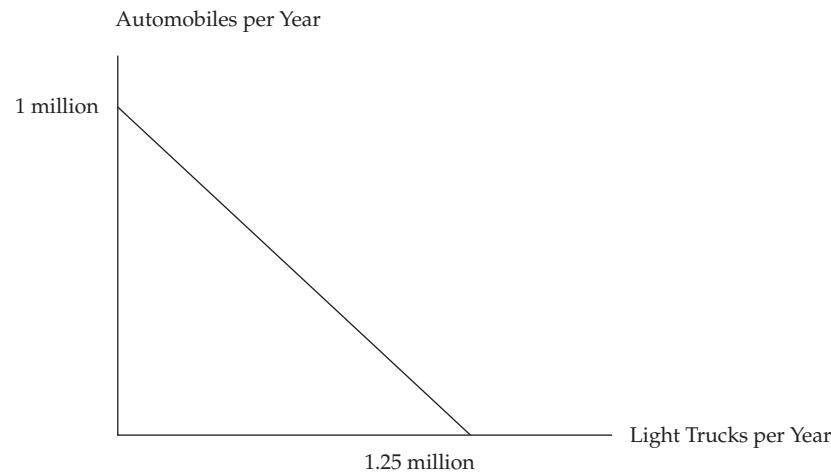
We can choose to measure either commodity on the vertical axis, so we arbitrarily choose cod. Note that if Nigel spends his entire budget on cod, he could buy 8kg. On the other hand, if he chooses to spend the entire budget on lamb, he could buy 12.8kg. Of course, he could spread his £128 between the two goods in any proportions he chooses, so the budget constraint is drawn as follows:

**Solution to 3:**

With quantity of cod measured on the vertical axis, the slope is equal to $-(P_L/P_C) = -10/16 = -0.625$. (Note: If we had chosen to measure quantity of lamb on the vertical axis, the slope would be inverted: $-(P_C/P_L) = -1.6$.)

4.2 The Production Opportunity Set

Companies face constraints on their production opportunities, just as consumers face limits on the bundles of goods that they can consume. Consider a company that produces two products using the same production capacity. For example, an automobile company might use the same factory to produce either automobiles or light trucks. If so, then the company is constrained by the limited capacity to produce vehicles. If it produces more trucks, it must reduce its production of automobiles; likewise, if it produces more automobiles, it must produce fewer trucks. The company's **production opportunity frontier** shows the maximum number of units of one good it can produce, for any given number of the other good that it chooses to manufacture. Such a frontier for the vehicle company might look something like that in Exhibit 9.

Exhibit 9 The Production Opportunity Frontier


Note: The production opportunity frontier for a vehicle manufacturer shows the maximum number of autos for any given level of truck production. In this example, the opportunity cost of a truck is 0.8 autos.

There are two important things to notice about this example. First, if the company devoted its entire production facility to the manufacture of automobiles, it could produce 1 million in a year. Alternatively, if it devoted its entire plant to trucks, it could produce 1.25 million a year. Of course, it could devote only part of the year's production to trucks, in which case it could produce automobiles during the remainder of the year. In this simple example, for every additional truck the company chooses to make, it would have to produce 0.8 fewer cars. That is, the **opportunity cost** of a truck is 0.8 cars, or the opportunity cost of a car is 1.25 trucks. The opportunity cost of trucks is the negative of the slope of the production opportunity frontier: $1/1.25$. And of course, the opportunity cost of an automobile is the inverse of that ratio, or 1.25.

The other thing to notice about this exhibit is that it assumes the opportunity cost of a truck is independent of how many trucks (and cars) the company produces. The production opportunity frontier is linear with a constant slope. Perhaps a more realistic example would be to increase marginal opportunity cost. As more and more trucks are produced, fewer inputs that are particularly well suited to producing truck inputs could be transferred to assist in their manufacture, causing the cost of trucks (in terms of cars) to rise as more trucks are produced. In this event, the production opportunity frontier would become steeper as the company moved its production point away from cars and toward more trucks, resulting in a frontier that would be concave as viewed from the origin.

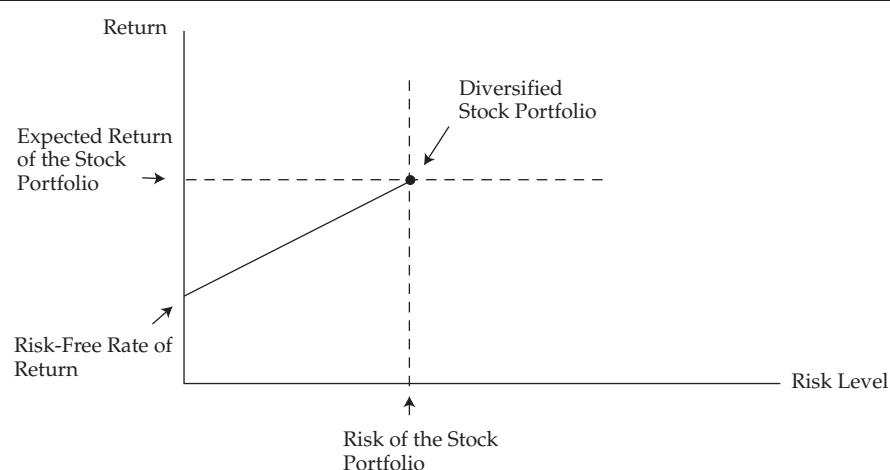
4.3 The Investment Opportunity Set

The investment opportunity set is examined in detail in readings on investments, but it is appropriate to examine it briefly here because we are learning about constraints on behavior. Consider possible investments in which one option might be to invest in an essentially risk-free asset, such as a US Treasury bill. There is virtually no possibility that the US government would default on a 90-day obligation to pay back an investor's purchase price, plus interest. Alternatively, an investor could put her money into a broadly diversified index of common shares. This investment will necessarily be more risky because of the fact that share prices fluctuate. If investors inherently find risk distasteful, then they will be reluctant to invest in a risky asset unless they expect to

receive, on average, a higher rate of return. Hence, it is reasonable to expect that a broadly diversified index of common shares will have an expected return that exceeds that of the risk-free asset, or else no one would hold that portfolio.

Our hypothetical investor could choose to put some of her funds in the risk-free asset and the rest in the common shares index. For each additional dollar invested in the common shares index, she can expect to receive a higher return, though not with certainty; so, she is exposing herself to more risk in the pursuit of a higher return. We can structure her investment opportunities as a frontier that shows the highest expected return consistent with any given level of risk, as shown in Exhibit 10. The investor's choice of a portfolio on the frontier will depend on her level of risk aversion.

Exhibit 10 The Investment Opportunity Frontier



Note: The investment opportunity frontier shows that as the investor chooses to invest a greater proportion of assets in the market portfolio, she can expect a higher return but also higher risk.

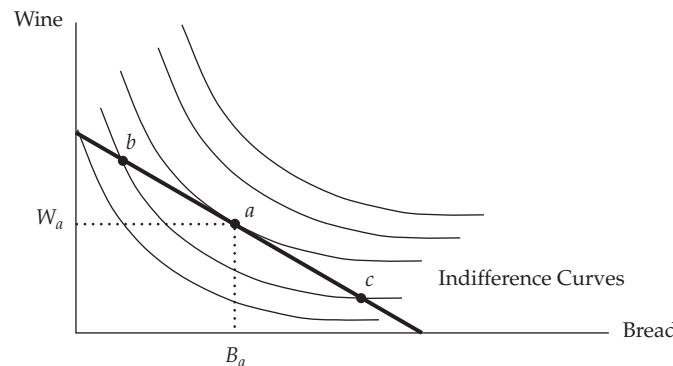
CONSUMER EQUILIBRIUM: MAXIMIZING UTILITY SUBJECT TO THE BUDGET CONSTRAINT

5

It would be wonderful if we could all consume as much of everything as we wanted, but unfortunately, most of us are constrained by income and prices. We now superimpose the budget constraint onto the preference map to model the actual choice of our consumer. This is a constrained (by the resources available to pay for consumption) optimization problem that every consumer must solve: Choose the bundle of goods and services that gets us as high on our ranking as possible, while not exceeding our budget.

5.1 Determining the Consumer's Equilibrium Bundle of Goods

In general, the consumer's constrained optimization problem consists of maximizing utility, subject to the budget constraint. If, for simplicity, we assume there are only two goods, wine and bread, then the problem appears graphically as in Exhibit 11.

Exhibit 11 Consumer Equilibrium


Note: Consumer equilibrium is achieved at point *a*, where the highest indifference curve is attained while not violating the budget constraint.

The consumer desires to reach the indifference curve that is farthest from the origin, while not violating the budget constraint. In this case, that pursuit ends at point *a*, where she is purchasing W_a ounces of wine along with B_a slices of bread per month. It is important to note that this equilibrium point represents the tangency between the highest indifference curve and the budget constraint. At a tangency point, the two curves have the same slope, meaning that the MRS_{BW} must be equal to the price ratio, P_B/P_W . Recall that the marginal rate of substitution is the rate at which the consumer is just *willing* to sacrifice wine for bread. Additionally, the price ratio is the rate at which she *must* sacrifice wine for another slice of bread. So, at equilibrium, the consumer is just willing to pay the opportunity cost that she must pay.

In contrast, consider another affordable bundle represented by point *b*. Certainly, the consumer is able to purchase that bundle because it lies on her budget constraint. However, the MRS_{BW} at that point is greater than the price ratio, meaning that she is *willing* to give up wine to obtain bread at a rate greater than she *must*. Hence, she will be better off moving downward along the budget constraint until she reaches the tangent point at *a*. In effect, she is willing to pay a higher price than she must for each additional unit of bread until she reaches B_a . For all of the units that she consumes up to B_a , we could say that the consumer is receiving consumer surplus, a concept we visited above when discussing the demand curve. Importantly, she would not purchase slices of bread beyond B_a at these prices because at a point like *c*, the marginal rate of substitution is less than the price ratio—meaning that the price for that additional unit is above her willingness to pay. Even though she could afford bundle *c*, it would not be the best use of her income.

EXAMPLE 5
Consumer Equilibrium

Currently, a consumer is buying both sorbet and gelato each week. His MRS_{GS} [marginal rate of substitution of gelato (*G*) for sorbet (*S*)] equals 0.75. The price of gelato is €1 per scoop, and the price of sorbet is €1.25 per scoop.

- 1 Determine whether the consumer is currently optimizing his budget over these two desserts. Justify your answer.
- 2 Explain whether the consumer should buy more sorbet or more gelato, given that he is not currently optimizing his budget.

Solution to 1:

In this example, the condition for consumer equilibrium is $MRS_{GS} = P_G/P_S$. Because $P_G/P_S = 0.8$ and $MRS_{GS} = 0.75$, the consumer is clearly not allocating his budget in a way that maximizes his utility, subject to his budget constraint.

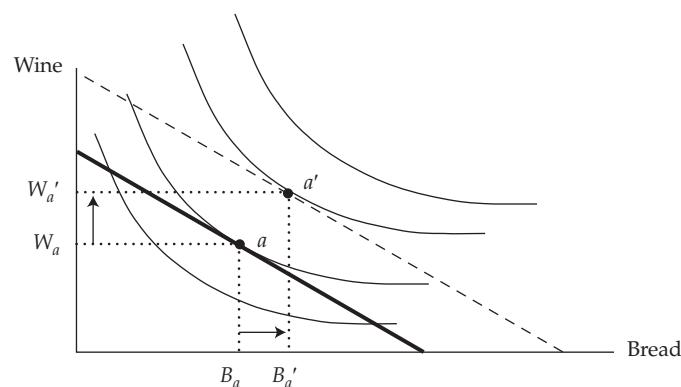
Solution to 2:

The MRS_{GS} is the rate at which the consumer is willing to give up sorbet to gain a small additional amount of gelato, which is 0.75 scoops of sorbet to gain one scoop of gelato. The price ratio, P_G/P_S (0.8), is the rate at which he *must* give up sorbet to gain an additional small amount of gelato. In this case, the consumer would be better off spending a little less on gelato and a little more on sorbet.

5.2 Consumer Response to Changes in Income: Normal and Inferior Goods

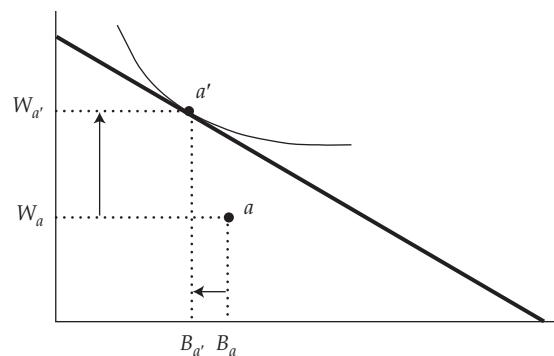
The consumer's behavior is constrained by his income and the prices he must pay for the goods he consumes. Consequently, if one or more of those parameters changes, the consumer is likely to change his consumption behavior. We first consider an increase in income. Recall from Exhibit 7, Panel C, that an increase in income simply shifts the budget constraint outward from the origin, parallel to itself. Exhibit 12 indicates such a shift and shows how the consumer would respond, in this case, by buying more of both bread and wine.

Exhibit 12 The Effect of an Increase in Income on a Normal Good



Note: The effect of an increase in income when both goods are normal is to increase the consumption of both.

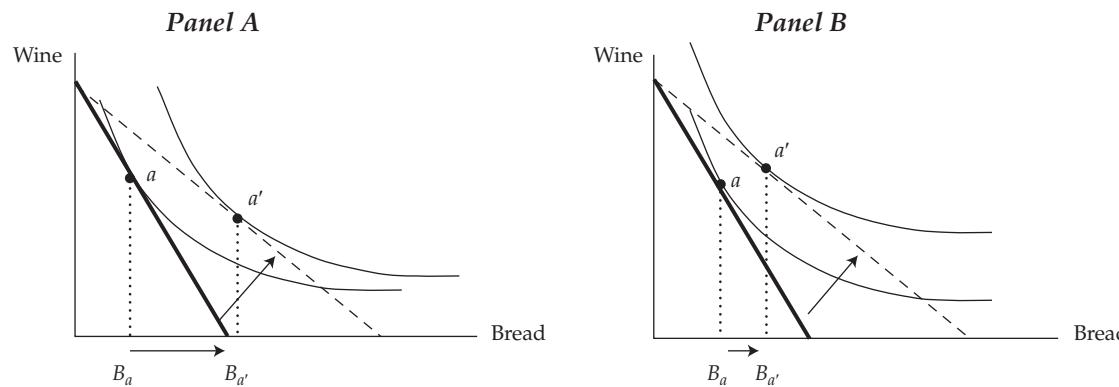
As we discovered, there is no restriction that the purchase of every good must respond to an increase in income with an increase in quantity. There, we defined *normal goods* as those with a positive response to an increase in income and *inferior goods* as those with a negative response to an increase in income. Suppose that bread is an inferior good for a particular consumer, whereas wine is a normal good. Exhibit 13 shows this consumer's purchase behavior when income increases. As income rises, the consumer purchases less bread but more wine.

Exhibit 13 The Effect of an Increase in Income on an Inferior Good

Note: The effect of an increase in income on the purchase of bread, an inferior good in this example, reduces the consumption of that good.

5.3 How the Consumer Responds to Changes in Price

We now hold income and the price of one good (wine) constant but decrease the price of the other good (bread). Recall that a decrease in the price of bread pivots the budget constraint outward along the horizontal axis but leaves the vertical intercept unchanged—as in Exhibit 14, where we examine two responses to the decrease in the price of bread.

Exhibit 14 Elastic and Inelastic Responses to a Decrease in Price

Note: When the price of bread falls, this consumer moves from bundle a to bundle a' , showing an elastic response.

Note: When the price of bread falls, this consumer moves from bundle a to bundle a' , showing an inelastic response.

In both cases, when the price of bread falls, the consumer buys more bread. But in the first case, he is quite responsive to the price change, responding with an elastic demand for bread. In the second case, the consumer is still responsive but much less so than the first consumer; this consumer's response to the price change is inelastic.

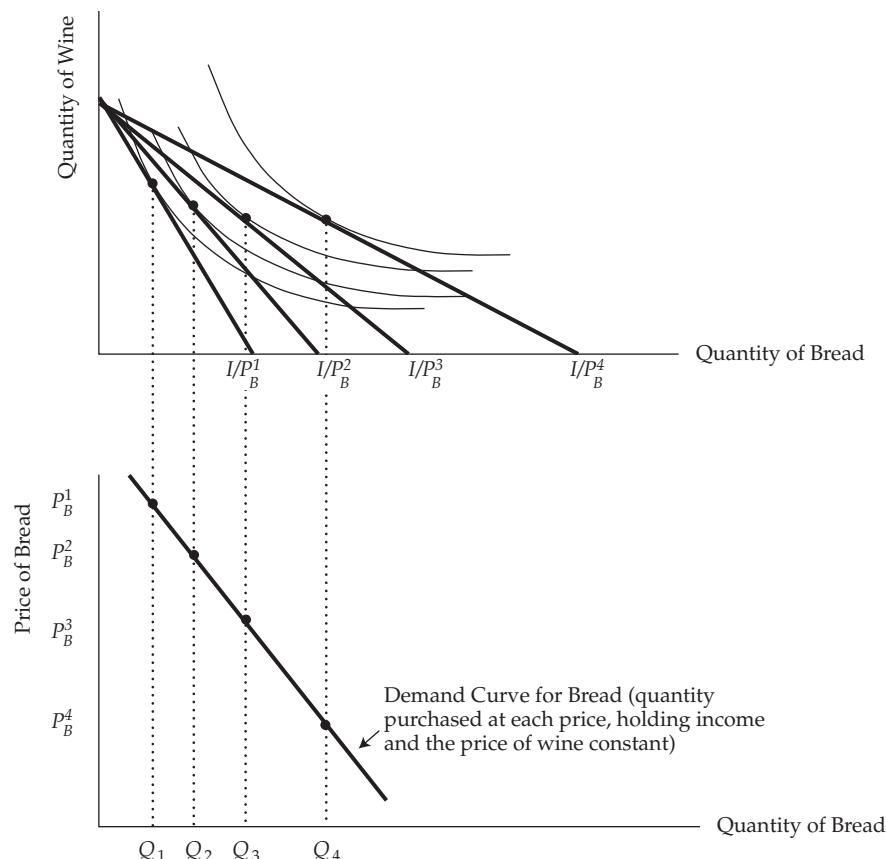
REVISITING THE CONSUMER'S DEMAND FUNCTION

6

We have now come to the reason why we wanted to explore consumer theory in the first place: We want to have a sound theoretical foundation for our use of consumer demand curves. Although we could merely assume that a consumer has a demand curve, we derive a richer understanding of that curve if we start with a more fundamental recognition of the consumer's preferences and her response to changes in the parameters that constrain her behavior in the marketplace.

6.1 Consumer's Demand Curve from Preferences and Budget Constraints

Recall that to draw a consumer's demand curve, we appealed to the assumption of "holding all other things constant" and held preferences, income, and the prices of all but one good constant. Graphically, we show such an exercise by representing a given utility function with a set of indifference curves, and then we superimpose a set of budget constraints, each one representing a different price of one of the goods. Exhibit 15 shows the result of this exercise. Notice that we are "stacking" two exhibits vertically to show both the indifference curves and budget constraints and the demand curve below them. In the upper exhibit, we have rotated the budget constraint rightward, indicating successively lower prices of bread, $P_B^1, P_B^2, P_B^3, P_B^4$, while holding income constant at I .

Exhibit 15 Deriving a Demand Curve

Note: A demand curve for bread is derived from the indifference curve map and a set of budget constraints representing different prices of bread.

This pair of diagrams deserves careful inspection. Notice first that the vertical axes are not the same. In the upper diagram, we represent the quantity of the *other* good, wine, whose price is being held constant, along with income. Hence, the budget constraints all have the same vertical intercept. But the price of bread is falling as we observe ever less steep budget constraints with horizontal intercepts moving rightward. Confronted with each respective budget constraint, the consumer finds the tangent point as indicated. This point corresponds to the respective quantities of bread, Q_1 , Q_2 , Q_3 , and Q_4 . Note also that the horizontal axes of the two diagrams are identical. They measure the quantity of bread purchased. Importantly, the vertical axis in the lower diagram measures the price of bread. As the price of bread falls, this consumer chooses to buy ever greater quantities, as indicated. The price–quantity combinations that result trace out this consumer's demand curve for bread in the lower diagram. For each tangent point in the upper diagram, there is a corresponding point in the lower diagram, tracing out the demand curve for bread.

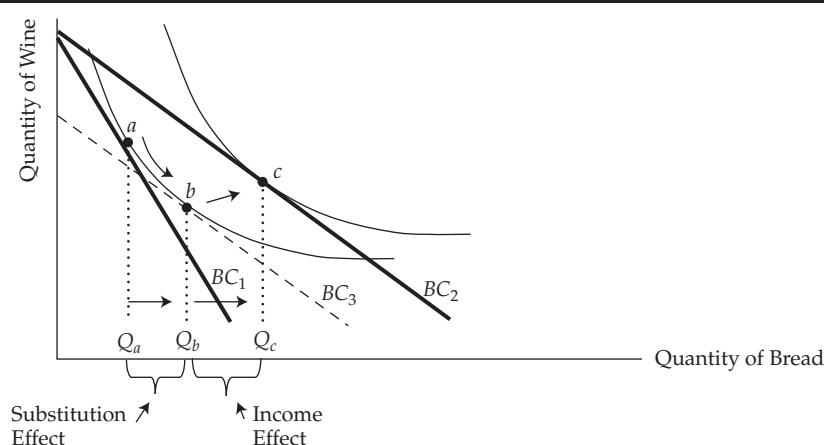
6.2 Substitution and Income Effects for a Normal Good

The law of demand says that when price falls, quantity demanded rises; however, it doesn't say why. We can answer that question by delving a little more deeply into consumer theory. When there is a decline in the price of a good that the consumer has been buying, two things happen. The good now becomes *relatively* less costly as

compared to other goods. That is, it becomes more of a bargain than other things the consumer could purchase; thus, more of this good gets substituted for other goods in the consumer's market basket. Additionally, though, with the decline in that price, the consumer's **real income** rises. We're not saying that the size of the consumer's paycheck changes; we're saying that the amount of goods that can be purchased with the same amount of money has increased. If this good is a normal good, then increases in income lead to increased purchases of this good. So the consumer tends to buy more when price falls for both reasons: the substitution effect and the income effect of a change in the price of a good.

A close look at indifference curves and budget constraints can demonstrate how these effects can be separated. Consider Exhibit 16, where we analyze Warren's response to a decrease in the price of bread. When the price of bread falls, as indicated by the pivoting in budget constraints from BC_1 to BC_2 , Warren buys more bread, increasing his quantity from Q_a to Q_c . That is the net effect of both the substitution effect and the income effect. We can see the partial impact of each of these effects by engaging in a mental exercise. Part of Warren's response is because of his increase in real income. We can remove that effect by subtracting some income from him, while leaving the new lower price in place. The dashed budget constraint shows the reduction in income that would be just sufficient to move Warren back to his original indifference curve. Notice that we are moving BC_2 inward, parallel to itself until it becomes just tangent to Warren's original indifference curve at point b . The price decrease was a good thing for Warren. An offsetting bad thing would be an income reduction. If the income reduction is just sufficient to leave Warren as well off but no better off than before the price change, then we have effectively removed the real income effect of the decrease in price. What's left of his response must be due to the pure substitution effect alone. So, we say that the substitution effect is shown by the move from point a to point b . If his income reduction were then restored, the resulting movement from point b to point c must be the pure income effect.

Exhibit 16 Substitution and Income Effects for a Normal Good



Note: Substitution effect (Q_a to Q_b) and the income effect (Q_b to Q_c) of a decrease in the price of a normal good.

An important thing to notice is that the pure substitution effect must always be in the direction of purchasing more when the price falls and purchasing less when the price rises. This is because of the diminishing marginal rate of substitution, or the convexity of the indifference curve. Look again at Exhibit 16. Note that the substitution effect is the result of changing from budget constraint 1 to budget constraint 3—or

moving the budget constraint along the original indifference curve, while maintaining tangency. Note that in the process, the budget constraint becomes less steep, just as the marginal rate of substitution decreases. Warren is no better off than before the changes, but his behavior has changed: He now buys more bread and less wine than before the offsetting changes in income and price. This reason for negatively sloped demand curves never changes.

Sellers can sometimes use income and substitution effects to their advantage. Think of something you often buy, perhaps lunch at your favorite café. How much would you be willing to pay for a “lunch club membership card” that would allow you to purchase lunches at, say, half price? If the café could extract from you the maximum amount each month that you would be willing to pay for the half-price option, then it would successfully have removed the income effect from you in the form of a monthly fixed fee. Notice that Exhibit 16 implies that you would end up buying more lunches each month than before you purchased the discount card, even though you would be no better or worse off than before. This is a way that sellers are sometimes able to extract consumer surplus by means of creative pricing schemes. It’s a common practice among big box retailers, sports clubs, and other users of what is called “two-part tariff pricing.”

EXAMPLE 6

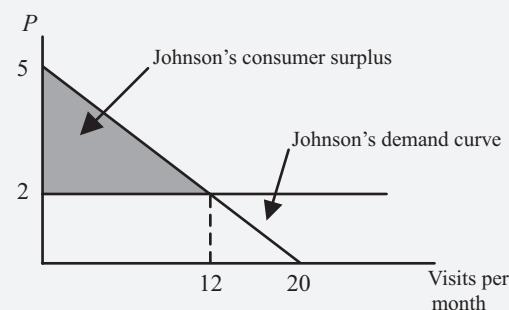
Two-Part Tariff Pricing

Nicole Johnson’s monthly demand for visits to her health club is given by the following equation: $Q^d = 20 - 4P$, where Q^d is visits per month and P is euros per visit. The health club’s marginal cost is fixed at €2 per visit.

- 1 Determine Johnson’s demand curve for health club visits per month.
- 2 Calculate how many visits Johnson would make per month if the club charged a price per visit equal to its marginal cost.
- 3 Calculate Johnson’s consumer surplus at the price determined in Question 2.
- 4 Calculate how much the club could charge Johnson each month for a membership fee.

Solution to 1:

$Q^d = 20 - 4P$, so when $P = 0$, $Q^d = 20$. Inverting, $P = 5 - 0.25Q$, so when $Q = 0$, $P = 5$.



Solution to 2:

$Q^d = 20 - 4(2) = 12$. Johnson would make 12 visits per month at a price of €2 per visit.

Solution to 3:

Johnson's consumer surplus can be measured as the area under her demand curve and above the price she pays, for a total of 12 visits: $CS = (\frac{1}{2})(12)(3) = 18$. Johnson would enjoy €18 per month consumer surplus.

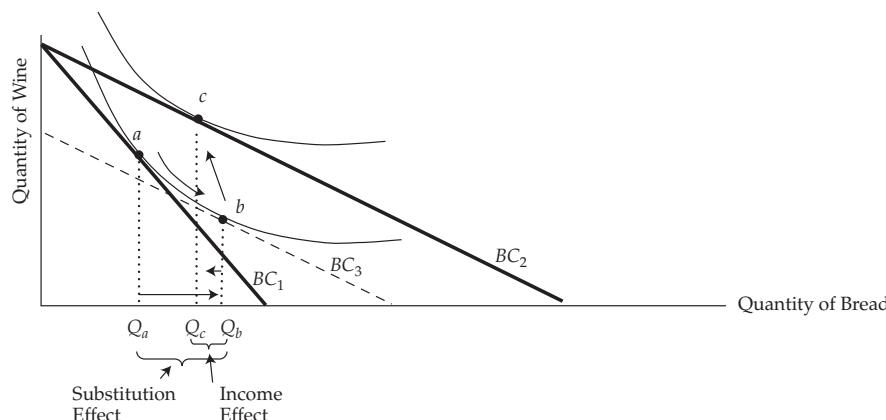
Solution to 4:

The club could extract all of Johnson's consumer surplus by charging her a monthly membership fee of €18 plus a per-visit price of €2. This is called a two-part tariff because it assesses one price per unit of the item purchased plus a per-month fee (sometimes called an "entry fee") equal to the buyer's consumer surplus evaluated at the per-unit price.

6.3 Income and Substitution Effects for an Inferior Good

We know that for some consumers and some goods, an increase in income leads to a decrease in the quantity purchased at each price. These goods are called *inferior* goods, and they have negative income elasticity of demand. When price falls, these goods still exhibit substitution and income effects, but they are in opposite directions. Consider Exhibit 17, in which we see a fairly standard set of indifference curves and budget constraints. But in this case, bread is an inferior good.

Exhibit 17 Income and Substitution Effects for an Inferior Good



Note: The income effect of a price decrease for an inferior good is opposite to the substitution effect, tending to mitigate the change in quantity.

Notice that when the bread's price falls, as indicated by the shift from budget constraint 1 (BC_1) to budget constraint 2 (BC_2), the consumer buys more bread, just as we would expect. That is, the consumer's demand curve is still negatively sloped. When we apply the income adjustment to isolate substitution effect from income effect, we shift the budget constraint back to constraint 3, reducing income sufficiently to place the consumer back on the original indifference curve. As before, the substitution effect is shown as a movement along the original indifference curve from point a to point b . The income effect is, as before, a movement from one indifference curve to the other, as shown by the movement from point b to point c . In this case, however, the income effect partially offsets the substitution effect, causing demand to be less elastic than if the two effects reinforced each other.

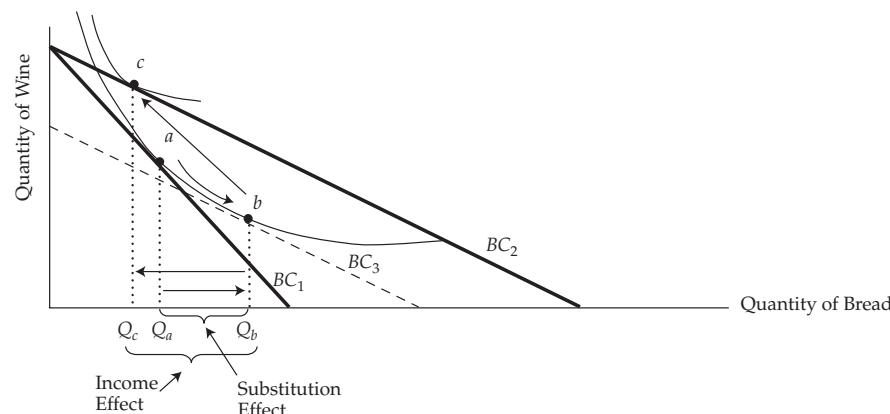
We see that for inferior goods, the income effect and the substitution effect are in opposite directions: The decrease in price causes the consumer to buy more, but the income effect tends to mitigate that effect. It's still true that a decrease in the price of bread represents an increase in real income. But in the case of an inferior good, the increased income causes the consumer to want to buy less of the good, not more. As long as the income effect has a lower magnitude than the substitution effect, the consumer still ends up buying more at the lower price. However, she buys a little less than she would if the good were normal. It is possible, though highly unlikely, for the income effect to have greater magnitude than the substitution effect. We examine that case next.

In the case of savings, the same type of effects can apply. For example, say interest rates rise. Individuals may save more because the reward (price) for saving has risen, and individuals substitute future consumption for present consumption. However, higher interest rates also imply that less saving is required to attain a given future amount of money. If the latter effect (the income effect) dominates, then it is possible to observe higher interest rates resulting in less savings.

6.4 Negative Income Effect Larger than Substitution Effect: Giffen Goods

In theory, it is possible for the income effect to be so strong and so negative as to overpower the substitution effect. If that were to occur, then a decrease in price could result in a *decrease* in quantity demanded and a *positively sloped* individual demand curve. Let us explore this curiosity in Exhibit 18.

Exhibit 18 Income and Substitution Effects for a Giffen Good



Note: Income and substitution effects of a fall in price for a Giffen good: When price declines, the consumer chooses to buy less of the good.

Once again, we decrease the price of bread as indicated by the pivoting of the budget constraint from BC_1 to BC_2 , and then we move the budget constraint parallel to itself leftward until it just touches the original indifference curve at point b to remove the income effect. What is left is the substitution effect. As always, the substitution effect causes the consumer to substitute more bread for less wine in the basket, as indicated by the movement along the indifference curve from point a to point b . But notice the odd result when we “give back” the income and move from BC_3 to BC_2 . The income effect for this inferior good (from point b to point c) is once again opposite in direction to the substitution effect, as is true for all inferior goods. But in this curious case, its

magnitude overwhelms the substitution effect: Point *c* lies *to the left* of point *a*. The consumer actually buys less of the good when its price falls, resulting in a positively sloped demand curve. If we reversed the analysis and increased the price of bread, this consumer would buy more bread when its price rose. Those inferior goods whose income effect is negative and greater in magnitude than the substitution effect are known as **Giffen goods**. Importantly, all Giffen goods must be inferior, but not all inferior goods are Giffen goods.

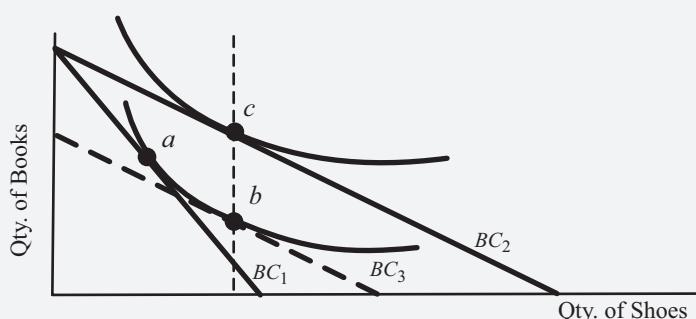
This curious result is originally attributed to Robert Giffen, who suspected that Irish peasants might have responded this way to increased prices of staples during the Irish potato famine in the nineteenth century. He reasoned that staples, such as potatoes, comprised a very large portion of the peasants' total budget. Additionally, potatoes could be a very inferior good, which simply means that when incomes fell, the peasants bought a lot more potatoes; and when incomes rose, they bought a lot fewer. Now because potatoes took up such a large part of total expenditures, any increase in the price of potatoes would result in a very substantial decrease in real income. This combination of strong inferiority coupled with a large amount of the budget spent on potatoes could, in theory, result in the negative income effect not only being opposite in direction to the substitution effect, but in fact overwhelming it.

Although some empirical studies have suggested the existence of Giffen goods, even if they existed they would be extraordinarily rare. Moreover, although they might exist for some small subset of consumers, it is highly unlikely that consumers as a whole would behave this way. So Giffen goods' role in microeconomics is greater than their role in the empirical world. True, they result in a positively sloped demand curve and they do not violate any of the axioms of consumer choice theory. But any company's manager who believes that if she raises the price of her product she will sell *more* of it is very likely to be disappointed.

EXAMPLE 7

Income and Substitution Effects of a Decrease in Price

Consider the following diagram of budget constraints and indifference curves for a consumer choosing to allocate her budget between books and shoes. Determine whether shoes are normal, inferior, or Giffen goods for this consumer.



Solution:

When the price of shoes falls, the original budget constraint pivots from BC_1 to BC_2 . The original tangent point was at *a* and is now at *c*. We can separate the substitution effect from the income effect by removing enough income to put the consumer back on the old indifference curve at point *b*. This is shown as a shift in the budget constraint from BC_2 to BC_3 , a parallel shift. The substitution effect is, therefore, from point *a* to point *b*, along the original indifference curve. The income effect is from point *b* to point *c*, but note that those two points are

on the same vertical line. In this case, shoes are on a borderline between normal and inferior. There is zero income effect. (If point *c* had been to the right of point *b*, shoes would be normal. And if *c* had been to the left of point *b*, they would be inferior. Finally, if *c* had been to the left of point *a*, shoes would be a Giffen good.)

6.5 Veblen Goods: Another Possibility for a Positively Sloped Demand Curve

Standard choice theory assumes that the consumer can always make comparisons among all pairs of bundles of goods and identify preferences before knowing anything about the prices of those baskets. Then, as we've seen, the consumer is constrained by income and prices and makes actual choices of which bundles to purchase with limited income. It is important to note that those preferences are assumed to exist even before knowing the prices at which those goods could be purchased. It is possible, however, that an item's price tag *itself* might help determine the consumer's preferences for it. Thorstein Veblen posited just such a circumstance in his concept of **conspicuous consumption**. According to this way of thinking, a consumer might derive utility out of being known by others to consume a so-called *high status* good, such as a luxury automobile or a very expensive piece of jewelry. Importantly, it is the high price itself that partly imparts value to such a good. If that is the case, then a consumer would actually value a good more if it had a higher price. So, it is argued that by increasing the price of a **Veblen good**, the consumer would be *more inclined* to purchase it, not less. In the extreme, it could be argued that the consumer's demand for such a good could be positively sloped, though this need not necessarily follow. In fact, of course, if any seller actually faced a positively sloped demand curve for her product, the rational response would be to increase price because she would sell more at the higher price. Ultimately, at some very high price, demand would necessarily become negatively sloped.

It is important to recognize that, although Veblen goods and Giffen goods share some characteristics, they are in fact quite different. Whether or not they actually exist, Giffen goods are not inconsistent with the fundamental axioms of demand theory. True, they would result in a violation of the law of demand, but that law is not a logical necessity. It is simply recognition that in virtually all observed cases, demand curves are negatively sloped. Giffen goods certainly would not be considered examples of "status goods," because an increase in income alone would result in a reduced interest in purchasing them. Veblen goods, on the other hand, derive their value from the ostentatious consumption of them as symbols of the purchaser's high status in society. If they exist, they are certainly not inferior goods. And they do violate the axioms of choice that form the foundation of accepted demand theory.

SUMMARY

This reading has explored how consumer preferences over baskets of goods and budget constraints translate into the demand curves posited by the demand and supply model of markets. Among the major points made are the following.

- Consumer choice theory is the branch of microeconomics that relates consumer demand curves to consumer preferences. Utility theory is a quantitative model of consumer preferences and is based on a set of axioms (assumptions that are assumed to be true). If consumer preferences are complete, transitive,

and insatiable, those preferences can be represented by an ordinal utility function and depicted by a set of indifference curves that are generally negatively sloped, convex from below, and do not cross for a given consumer.

- A consumer's relative strength of preferences can be inferred from his marginal rate of substitution of good X for good Y (MRS_{XY}), which is the rate at which the consumer is willing to sacrifice good Y to obtain an additional small increment of good X . If two consumers have different marginal rates of substitution, they can both benefit from the voluntary exchange of one good for the other.
- A consumer's attainable consumption options are determined by her income and the prices of the goods she must purchase to consume. The set of options available is bounded by the budget constraint, a negatively sloped linear relationship that shows the highest quantity of one good that can be purchased for any given amount of the other good being bought.
- Analogous to the consumer's consumption opportunity set are, respectively, the production opportunity set and the investment opportunity set. A company's production opportunity set represents the greatest quantity of one product that a company can produce, for any given amount of the other good it produces. The investment opportunity set represents the highest return an investor can expect, for any given amount of risk undertaken.
- Consumer equilibrium is obtained when utility is maximized, subject to the budget constraint, generally depicted as a tangency between the highest attainable indifference curve and the fixed budget constraint. At that tangency, the MRS_{XY} is just equal to the two goods' price ratio, P_X/P_Y —or that bundle such that the rate at which the consumer is just willing to sacrifice good Y for good X is equal to the rate at which, based on prices, she must sacrifice good Y for good X .
- If the consumer's income and the price of all other goods are held constant and the price of good X is varied, the set of consumer equilibria that results will yield that consumer's demand curve for good X . In general, we expect the demand curve to have a negative slope (the law of demand) because of two influences: income and substitution effects of a decrease in price. Normal goods have a negatively sloped demand curve. For normal goods, income and substitution effects reinforce one another. However, for inferior goods, the income effect offsets part or all of the substitution effect. In the case of the Giffen good, the income effect of this very inferior good overwhelms the substitution effect, resulting in a positively sloped demand curve.
- In accepted microeconomic consumer theory, the consumer is assumed to be able to judge the value of any given bundle of goods without knowing anything about their prices. Then, constrained by income and prices, the consumer is assumed to be able to choose the optimal bundle of goods that is in the set of available options. It is possible to conceive of a situation in which the consumer cannot truly value a good until the price is known. In these Veblen goods, the price is used by the consumer to signal the consumer's status in society. Thus, to some extent, the higher the price of the good, the more value it offers to the consumer. In the extreme case, this could possibly result in a positively sloped demand curve. This result is similar to a Giffen good, but the two goods are fundamentally different.

PRACTICE PROBLEMS

- 1 A child indicates that she prefers going to the zoo over the park and prefers going to the beach over the zoo. When given the choice between the park and the beach, she chooses the park. Which of the following assumptions of consumer preference theory is she *most likely* violating?
 - A Non-satiation.
 - B Complete preferences.
 - C Transitive preferences.
- 2 Which of the following ranking systems *best* describes consumer preferences within a utility function?
 - A Util.
 - B Ordinal.
 - C Cardinal.
- 3 Which of the following statements *best* explains why indifference curves are generally convex as viewed from the origin?
 - A The assumption of non-satiation results in convex indifference curves.
 - B The marginal rate of substitution of one good for another remains constant along an indifference curve.
 - C The marginal utility gained from one additional unit of a good versus another diminishes the more one has of the first good.
- 4 If a consumer's marginal rate of substitution of good X for good Y (MRS_{XY}) is equal to 2, then the:
 - A consumer is willing to give up 2 units of X for 1 unit of Y .
 - B slope of a line tangent to the indifference curve at that point is 2.
 - C slope of a line tangent to the indifference curve at that point is -2.
- 5 In the case of two goods, x and y , which of the following statements is *most likely* true? Maximum utility is achieved:
 - A along the highest indifference curve below the budget constraint line.
 - B at the tangency between the highest attainable indifference curve and the budget constraint line.
 - C when the marginal rate of substitution is equal to the ratio of the price of good y to the price of good x .
- 6 In the case of a normal good with a decrease in own price, which of the following statements is *most likely* true?
 - A Both the substitution and income effects lead to an increase in the quantity purchased.
 - B The substitution effect leads to an increase in the quantity purchased, while the income effect has no impact.
 - C The substitution effect leads to an increase in the quantity purchased, while the income effect leads to a decrease.
- 7 For a Giffen good, the:
 - A demand curve is positively sloped.
 - B substitution effect overwhelms the income effect.

- C income and substitution effects are in the same direction.
- 8 Which of the following statements *best* illustrates the difference between a Giffen good and a Veblen good?
- A The Giffen good alone is an inferior good.
 - B The substitution effect for each is in opposite directions.
 - C The Veblen good alone has a positively sloped demand curve.

SOLUTIONS

- 1 C is correct. If the child prefers the zoo over the park and the beach over the zoo, then she should prefer the beach over the park according to the axiom of transitive preferences.
- 2 B is correct. Utility functions only allow ordinal rankings of consumer preferences.
- 3 C is correct. The slope of the indifference curve at any point gives the marginal rate of substitution of one good for another. The curve is convex because the marginal value of one good versus another decreases the more one has of the first good.
- 4 C is correct. The marginal rate of substitution is equal to the negative of the slope of the tangent to the indifference curve at that point, or -2.
- 5 B is correct. Maximum utility is achieved where the highest attainable indifference curve intersects with just one point (the tangency) on the budget constraint line.
- 6 A is correct. In the case of normal goods, the income and substitution effects are reinforcing, leading to an increase in the amount purchased after a drop in price.
- 7 A is correct. The income effect overwhelms the substitution effect such that an increase in the price of the good results in greater demand for the good, resulting in a positively sloped demand curve.
- 8 A is correct. Veblen goods are not inferior goods, whereas Giffen goods are. An increase in income for consumers of a Veblen good leads to an increase in the quantity purchased at each price. The opposite is true for a Giffen good.

Glossary

Budget constraint A constraint on spending or investment imposed by wealth or income.

Complete preferences The assumption that a consumer is able to make a comparison between any two possible bundles of goods.

Conspicuous consumption Consumption of high status goods, such as a luxury automobile or a very expensive piece of jewelry.

Consumer choice theory The theory relating consumer demand curves to consumer preferences.

Consumption basket A specific combination of the goods and services that a consumer wants to consume.

Consumption bundle A specific combination of the goods and services that a consumer wants to consume.

Giffen good A good that is consumed more as the price of the good rises.

Income constraint The constraint on a consumer to spend, in total, no more than his income.

Indifference curve A curve representing all the combinations of two goods or attributes such that the consumer is entirely indifferent among them.

Indifference curve map A group or family of indifference curves, representing a consumer's entire utility function.

Marginal rate of substitution The rate at which one is willing to give up one good to obtain more of another.

Non-satiation The assumption that the consumer could never have so much of a preferred good that she would refuse any more, even if it were free; sometimes referred to as the "more is better" assumption.

Opportunity cost The value that investors forgo by choosing a particular course of action; the value of something in its best alternative use.

Production opportunity frontier Curve describing the maximum number of units of one good a company can produce, for any given number of the other good that it chooses to manufacture.

Real income Income adjusted for the effect of inflation on the purchasing power of money.

Transitive preferences The assumption that when comparing any three distinct bundles, A , B , and C , if A is preferred to B and simultaneously B is preferred to C , then it must be true that A is preferred to C .

Utility function A mathematical representation of the satisfaction derived from a consumption basket.

Utils A unit of utility.

Veblen good A good that increases in desirability with price.

READING

15

Demand and Supply Analysis: The Firm

by Gary L. Arbogast, CFA, and Richard V. Eastin, PhD

Gary L. Arbogast, CFA (USA). Richard V. Eastin, PhD, is at the University of Southern California (USA).

LEARNING OUTCOMES

Mastery	<i>The candidate should be able to:</i>
<input type="checkbox"/>	a. calculate, interpret, and compare accounting profit, economic profit, normal profit, and economic rent;
<input type="checkbox"/>	b. calculate and interpret and compare total, average, and marginal revenue;
<input type="checkbox"/>	c. describe a firm's factors of production;
<input type="checkbox"/>	d. calculate and interpret total, average, marginal, fixed, and variable costs;
<input type="checkbox"/>	e. determine and describe breakeven and shutdown points of production;
<input type="checkbox"/>	f. describe approaches to determining the profit-maximizing level of output;
<input type="checkbox"/>	g. describe how economies of scale and diseconomies of scale affect costs;
<input type="checkbox"/>	h. distinguish between short-run and long-run profit maximization;
<input type="checkbox"/>	i. distinguish among decreasing-cost, constant-cost, and increasing-cost industries and describe the long-run supply of each;
<input type="checkbox"/>	j. calculate and interpret total, marginal, and average product of labor;
<input type="checkbox"/>	k. describe the phenomenon of diminishing marginal returns and calculate and interpret the profit-maximizing utilization level of an input;
<input type="checkbox"/>	l. determine the optimal combination of resources that minimizes cost.

1

INTRODUCTION

In studying decision making by consumers and businesses, microeconomics gives rise to the theory of the consumer and theory of the firm as two branches of study.

The **theory of the consumer** is the study of consumption—the demand for goods and services—by utility-maximizing individuals. The **theory of the firm**, the subject of this reading, is the study of the supply of goods and services by profit-maximizing firms. Conceptually, profit is the difference between revenue and costs. Revenue is a function of selling price and quantity sold, which are determined by the demand and supply behavior in the markets into which the firm sells/provides its goods or services. Costs are a function of the demand and supply interactions in resource markets, such as markets for labor and for physical inputs. The main focus of this reading is the cost side of the profit equation for companies competing in market economies under perfect competition. A subsequent reading will examine the different types of markets into which a firm may sell its output.

The study of the profit-maximizing firm in a single time period is the essential starting point for the analysis of the economics of corporate decision making. Furthermore, with the attention given to earnings by market participants, the insights gained by this study should be practically relevant. Among the questions this reading will address are the following:

- How should profit be defined from the perspective of suppliers of capital to the firm?
- What is meant by factors of production?
- How are total, average, and marginal costs distinguished, and how is each related to the firm's profit?
- What roles do marginal quantities (selling prices and costs) play in optimization?

This reading is organized as follows: Section 2 discusses the types of profit measures, including what they have in common, how they differ, and their uses and definitions. Section 3 covers the revenue and cost inputs of the profit equation and the related topics of breakeven analysis, shutdown point of operation, market entry and exit, cost structures, and scale effects. In addition, the economic outcomes related to a firm's optimal supply behavior over the short run and long run are presented in this section. A summary and practice problems conclude the reading.

2

OBJECTIVES OF THE FIRM

This reading assumes that the objective of the firm is to maximize profit over the period ahead. Such analysis provides both tools (e.g., optimization) and concepts (e.g., productivity) that can be adapted to more-complex cases and also provides a set of results that may offer useful approximations in practice. The price at which a given quantity of a good can be bought or sold is assumed to be known with certainty (i.e., the theory of the firm under conditions of certainty). The main contrast of this type of analysis is to the theory of the firm under conditions of uncertainty, where prices, and therefore profit, are uncertain. Under market uncertainty, a range of possible profit outcomes is associated with the firm's decision to produce a given quantity of goods or services over a specific time period. Such complex theory typically makes simplifying assumptions. When managers of for-profit companies have been surveyed about the objectives of the companies they direct, researchers have often concluded that a)

companies frequently have multiple objectives; b) objectives can often be classified as focused on profitability (e.g., maximizing profits, increasing market share) or on controlling risk (e.g., survival, stable earnings growth); and c) managers in different countries may have different emphases.

Finance experts frequently reconcile profitability and risk objectives by stating that the objective of the firm is, or should be, **shareholder wealth maximization** (i.e., to maximize the market value of shareholders' equity). This theory states that firms try, or should try, to increase the wealth of their owners (shareholders) and that market prices balance returns against risk. However, complex corporate objectives may exist in practice. Many analysts view profitability as the single most important measure of business performance. Without profit, the business eventually fails; with profit, the business can survive, compete, and prosper. The question is: What is profit? Economists, accountants, investors, financial analysts, and regulators view profit from different perspectives. The starting point for anyone who is doing profit analysis is to have a solid grasp of how various forms of profit are defined and how to interpret the profit based on these different definitions.

By defining profit in general terms as the difference between total revenue and total costs, profit maximization involves the following expression:

$$\Pi = TR - TC \quad (1)$$

where Π is profit, TR is total revenue, and TC is total costs. TC can be defined as accounting costs or economic costs, depending on the objectives and requirements of the analyst for evaluating profit. The characteristics of the product market, where the firm sells its output or services, and of the resource market, where the firm purchases resources, play an important role in the determination of profit. Key variables that determine TC are the level of output, the firm's efficiency in producing that level of output when utilizing inputs, and resource prices as established by resource markets. TR is a function of output and product price as determined by the firm's product market.

2.1 Types of Profit Measures

The economics discipline has its own concept of profit, which differs substantially from what accountants consider profit. There are thus two basic types of profit—accounting and economic—and analysts need to be able to interpret each correctly and to understand how they are related to each other. In the theory of the firm, however, *profit* without further qualification refers to *economic profit*.

2.1.1 Accounting Profit

Accounting profit is generally defined as net income reported on the income statement according to standards established by private and public financial oversight bodies that determine the rules for financial reporting. One widely accepted definition of accounting profit—also known as net profit, net income, or net earnings—states that it equals revenue less all **accounting (or explicit) costs**. Accounting or explicit costs are payments to non-owner parties for services or resources that they supply to the firm. Often referred to as the “bottom line” (the last income figure in the income statement), accounting profit is what is left after paying all accounting costs—whether the expense is a cash outlay or not. When accounting profit is negative, it is called an **accounting loss**. Equation 2 summarizes the concept of accounting profit:

$$\text{Accounting profit} = \text{Total revenue} - \text{Total accounting costs} \quad (2)$$

When defining profit as accounting profit, the TC term in Equation 1 becomes total accounting costs, which include only the explicit costs of doing business. Let us consider two businesses: a start-up company and a publicly traded corporation. Suppose that for the start-up, total revenue in the business's first year is €3,500,000 and

total accounting costs are €3,200,000. Accounting profit is €3,500,000 – €3,200,000 = €300,000. The corresponding calculation for the publicly traded corporation, let us suppose, is \$50,000,000 – \$48,000,000 = \$2,000,000. Note that total accounting costs in either case include interest expense—which represents the return required by suppliers of debt capital—because interest expense is an explicit cost.

2.1.2 Economic Profit and Normal Profit

Economic profit (also known as **abnormal profit** or **supernormal profit**) may be defined broadly as accounting profit less the implicit opportunity costs not included in total accounting costs.

$$\text{Economic profit} = \text{Accounting profit} - \text{Total implicit opportunity costs} \quad (3a)$$

We can define a term, **economic cost**, equal to the sum of total accounting costs and implicit opportunity costs. Economic profit is therefore equivalently defined as:

$$\text{Economic profit} = \text{Total revenue} - \text{Total economic costs} \quad (3b)$$

For publicly traded corporations, the focus of investment analysts' work, the cost of equity capital is the largest and most readily identified implicit opportunity cost omitted in calculating total accounting cost. Consequently, economic profit can be defined for publicly traded corporations as accounting profit less the required return on equity capital.

Examples will make these concepts clearer. Consider the start-up company for which we calculated an accounting profit of €300,000 and suppose that the entrepreneurial executive who launched the start-up took a salary reduction of €100,000 per year relative to the job he left. That €100,000 is an opportunity cost of involving him in running the start-up. Besides labor, financial capital is a resource. Suppose that the executive, as sole owner, makes an investment of €1,500,000 to launch the enterprise and that he might otherwise expect to earn €200,000 per year on that amount in a similar risk investment. Total implicit opportunity costs are €100,000 + €200,000 = €300,000 per year and economic profit is zero: €300,000 – €300,000 = €0. For the publicly traded corporation, we consider the cost of equity capital as the only implicit opportunity cost identifiable. Suppose that equity investment is \$18,750,000 and shareholders' required rate of return is 8 percent so that the dollar cost of equity capital is \$1,500,000. Economic profit for the publicly traded corporation is therefore \$2,000,000 (accounting profit) less \$1,500,000 (cost of equity capital) or \$500,000.

For the start-up company, economic profit was zero. Total economic costs were just covered by revenues and the company was not earning a euro more nor less than the amount that meets the opportunity costs of the resources used in the business. Economists would say the company was earning a normal profit (economic profit of zero). In simple terms, **normal profit** is the level of accounting profit needed to just cover the implicit opportunity costs ignored in accounting costs. For the publicly traded corporation, normal profit was \$1,500,000: normal profit can be taken to be the cost of equity capital (in money terms) for such a company or the dollar return required on an equal investment by equity holders in an equivalently risky alternative investment opportunity. The publicly traded corporation actually earned \$500,000 in excess of normal profit, which should be reflected in the common shares' market price.

Thus, the following expression links accounting profit to economic profit and normal profit:

$$\text{Accounting profit} = \text{Economic profit} + \text{Normal profit} \quad (4)$$

When accounting profit equals normal profit, economic profit is zero. Further, when accounting profit is greater than normal profit, economic profit is positive; and when accounting profit is less than normal profit, economic profit is negative (the firm has an **economic loss**).

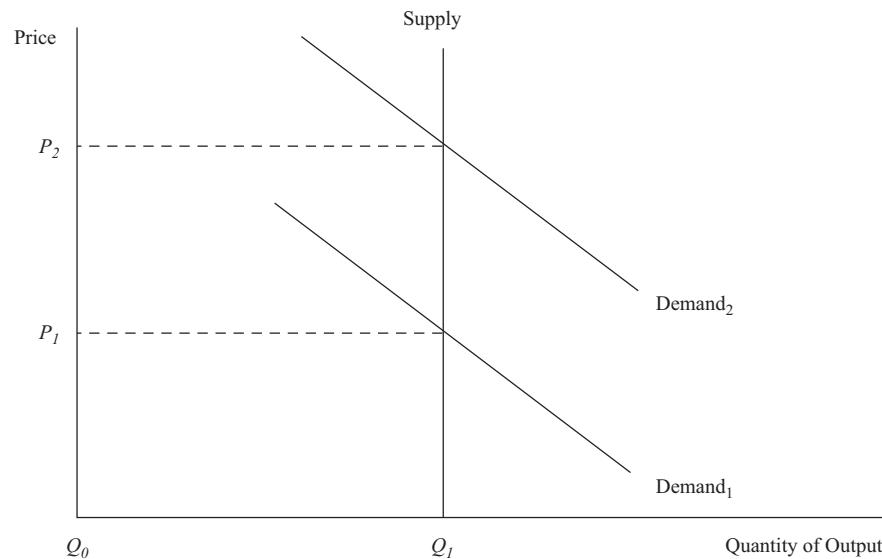
Economic profit for a firm can originate from sources such as:

- competitive advantage;
- exceptional managerial efficiency or skill;
- difficult to copy technology or innovation (e.g., patents, trademarks, and copyrights);
- exclusive access to less-expensive inputs;
- fixed supply of an output, commodity, or resource;
- preferential treatment under governmental policy;
- large increases in demand where supply is unable to respond fully over time;
- exertion of monopoly power (price control) in the market; and
- market barriers to entry that limit competition.

Any of the above factors may lead the firm to have positive net present value investment (NPV) opportunities. Access to positive NPV opportunities and therefore profit in excess of normal profits in the short run may or may not exist in the long run, depending on the potential strength of competition. In highly competitive market situations, firms tend to earn the normal profit level over time because ease of market entry allows for other competing firms to compete away any economic profit over the long run. Economic profit that exists over the long run is usually found where competitive conditions persistently are less than perfect in the market.

2.1.3 *Economic Rent*

The surplus value known as **economic rent** results when a particular resource or good is fixed in supply (with a vertical supply curve) and market price is higher than what is required to bring the resource or good onto the market and sustain its use. Essentially, demand determines the price level and the magnitude of economic rent that is forthcoming from the market. Exhibit 1 illustrates this concept, where P_1 is the price level that yields a normal profit return to the business that supplies the item. When demand increases from Demand₁ to Demand₂, price rises to P_2 , where at this higher price level economic rent is created. The amount of this economic rent is calculated as $(P_2 - P_1) \times Q_1$. The firm has not done anything internally to merit this special reward: It benefits from an increase in demand in conjunction with a supply curve that does not fully adjust with an increase in quantity when price rises.

Exhibit 1 Economic Rent

Because of their limited availability in nature, certain resources—such as land and specialty commodities—possess highly **inelastic supply** curves in both the short run and long run (shown in Exhibit 1 as a vertical supply curve). When supply is relatively inelastic, a high degree of market demand can result in pricing that creates economic rent. This economic rent results from the fact that when price increases, the quantity supplied does not change or, at the most, increases only slightly. This is because of the fixation of supply by nature or by such artificial constraints as government policy.

How is the concept of economic rent useful in financial analysis? Commodities or resources that command economic rent have the potential to reward equity investors more than what is required to attract their capital to that activity, resulting in greater shareholders' wealth. Evidence of economic rent attracts additional capital funds to the economic endeavor. This new investment capital increases shareholders' value as investors bid up share prices of existing firms. Any commodity, resource, or good that is fixed or nearly fixed in supply has the potential to yield economic rent. From an analytical perspective, one can obtain industry supply data to calculate the **elasticity of supply**, which measures the sensitivity of quantity supplied to a change in price. If quantity supplied is relatively unresponsive (**inelastic**) to price changes, then a potential condition exists in the market for economic rent. A reliable forecast of changes in demand can indicate the degree of any economic rent that is forthcoming from the market in the future. When one is analyzing fixed or nearly fixed supply markets (e.g., gold), a fundamental comprehension of demand determinants is necessary to make rational financial decisions based on potential economic rent.

EXAMPLE 1**Economic Rent and Investment Decision Making**

The following market data show the global demand, global supply, and price on an annual basis for gold over the period 2006–2008. Based on the data, what observation can be made about market demand, supply, and economic rent?

Year	2006	2007	2008	Percent Change 2006–2008
Supply (in metric tons)	3,569	3,475	3,508	-1.7
Demand (in metric tons)	3,423	3,552	3,805	+11.2
Average spot price (in US\$)	603.92	695.39	871.65	+44.3

Source: GFMS and World Gold Council.

Solution:

The amount of total gold supplied to the world market over this period has actually declined slightly by 1.7 percent during a period when there was a double-digit increase of 11.2 percent in demand. As a consequence, the spot price has dramatically increased by 44.3 percent. Economic rent has resulted from this market relationship of a relatively fixed supply of gold and a rising demand for it.

2.2 Comparison of Profit Measures

All three types of profit are interconnected because, according to Equation 4, accounting profit is the summation of normal and economic profit. In the short run, the normal profit rate is relatively stable, which makes accounting and economic profit the two variable terms in the profit equation. Over the longer term, all three types of profit are variable, where the normal profit rate can change according to investment returns across firms in the industry.

Normal profit is necessary to stay in business in the long run; positive economic profit is not. A business can survive indefinitely by just making the normal profit return for investors. Failing to earn normal profits over the long run has a debilitating impact on the firm's ability to access capital and to function properly as a business enterprise. Consequentially, the market value of equity and shareholders' wealth deteriorates whenever risk to achieving normal profit materializes and the firm fails to reward investors for their risk exposure and for the opportunity cost of their equity capital.

To summarize, the ultimate goal of analyzing the different types of profit is to determine how their relationships to one another influence the firm's market value of equity. Exhibit 2 compares accounting, normal, and economic profits in terms of how a firm's market value of equity is impacted by the relationships among the three types of profit.

Exhibit 2 Relationship of Accounting, Normal, and Economic Profit to Equity Value

Relationship between Accounting Profit and Normal Profit	Economic Profit	Firm's Market Value of Equity
Accounting profit > Normal profit	Economic profit > 0 and firm is able to protect economic profit over the long run	Positive effect
Accounting profit = Normal profit	Economic profit = 0	No effect
Accounting profit < Normal profit	Economic profit < 0 implies economic loss	Negative effect

3

ANALYSIS OF REVENUE, COSTS, AND PROFITS

To fully comprehend the dimensions of profit maximization, one must have a detailed understanding of the revenue and cost variables that determine profit.

Revenue and cost flows are calculated in terms of total, average, and marginal. A total is the summation of all individual components. For example, total cost is the summation of all costs that are incurred by the business. Total revenue is the sum of the revenues from all the business's units. In the theory of the firm, averages and marginals are calculated with respect to the quantity produced and sold in a single period (as opposed to averaging a quantity over a number of time periods). For example, average revenue is calculated by dividing total revenue by the number of items sold. To calculate a marginal term, take the change in the total and divide by the change in the quantity number.

Exhibit 3 shows a summary of the terminology and formulas pertaining to profit maximization, where profit is defined as total revenue minus total economic costs. Note that the definition of profit is the economic version, which recognizes that the implicit opportunity costs of equity capital, in addition to explicit accounting costs, are economic costs. The first main category consists of terms pertaining to the revenue side of the profit equation: total revenue, average revenue, and marginal revenue. Cost terms follow with an overview of the different types of costs—total, average, and marginal.

Exhibit 3 Summary of Profit, Revenue, and Cost Terms

Term	Calculation
Profit	
(Economic) profit	Total revenue minus total economic cost; $(TR - TC)$
Revenue	
Total revenue (TR)	Price times quantity ($P \times Q$), or the sum of individual units sold times their respective prices; $\sum(P_i \times Q_i)$
Average revenue (AR)	Total revenue divided by quantity; $(TR \div Q)$
Marginal revenue (MR)	Change in total revenue divided by change in quantity; $(\Delta TR \div \Delta Q)$
Costs	
Total fixed cost (TFC)	Sum of all fixed expenses; here defined to include all opportunity costs
Total variable cost (TVC)	Sum of all variable expenses, or per unit variable cost times quantity; (per unit $VC \times Q$)
Total costs (TC)	Total fixed cost plus total variable cost; $(TFC + TVC)$
Average fixed cost (AFC)	Total fixed cost divided by quantity; $(TFC \div Q)$
Average variable cost (AVC)	Total variable cost divided by quantity; $(TVC \div Q)$
Average total cost (ATC)	Total cost divided by quantity; $(TC \div Q)$ or $(AFC + AVC)$
Marginal cost (MC)	Change in total cost divided by change in quantity; $(\Delta TC \div \Delta Q)$

3.1 Profit Maximization

In free markets—and even in regulated market economies—profit maximization tends to promote economic welfare and a higher standard of living, and creates wealth for investors. Profit motivates businesses to use resources efficiently and to concentrate

on activities in which they have a competitive advantage. Most economists believe that profit maximization promotes allocational efficiency—that resources flow into their highest valued uses.

Overall, the functions of profit are as follows:

- Rewards entrepreneurs for risk taking when pursuing business ventures to satisfy consumer demand.
- Allocates resources to their most-efficient use; input factors flow from sectors with economic losses to sectors with economic profit, where profit reflects goods most desired by society.
- Spurs innovation and the development of new technology.
- Stimulates business investment and economic growth.

There are three approaches to calculate the point of profit maximization. First, given that profit is the difference between total revenue and total costs, maximum profit occurs at the output level where this difference is the greatest. Second, maximum profit can also be calculated by comparing revenue and cost for each individual unit of output that is produced and sold. A business increases profit through greater sales as long as per-unit revenue exceeds per-unit cost on the next unit of output sold. Profit maximization takes place at the point where the last individual output unit breaks even. Beyond this point, total profit decreases because the per-unit cost is higher than the per-unit revenue from successive output units. A third approach compares the revenue generated by each resource unit with the cost of that unit. Profit contribution occurs when the revenue from an input unit exceeds its cost. The point of profit maximization is reached when resource units no longer contribute to profit. All three approaches yield the same profit-maximizing quantity of output. (These approaches will be explained in greater detail later.)

Because profit is the difference between revenue and cost, an understanding of profit maximization requires that we examine both of those components. Revenue comes from the demand for the firm's products, and cost comes from the acquisition and utilization of the firm's inputs in the production of those products.

3.1.1 Total, Average, and Marginal Revenue

This section briefly examines demand and revenue in preparation for addressing cost. Unless the firm is a pure **monopolist** (i.e., the only seller in its market), there is a difference between market demand and the demand facing an individual firm. A later reading will devote much more time to understanding the various competitive environments (perfect competition, monopolistic competition, oligopoly, and monopoly), known as **market structure**. To keep the analysis simple at this point, we will note that competition could be either perfect or imperfect. In **perfect competition**, the individual firm has virtually no impact on market price, because it is assumed to be a very small seller among a very large number of firms selling essentially identical products. Such a firm is called a **price taker**. In the second case, the firm does have at least some control over the price at which it sells its product because it must lower its price to sell more units.

Exhibit 4 presents total, average, and marginal revenue data for a firm under the assumption that the firm is price taker at each relevant level of quantity of goods sold. Consequently, the individual seller faces a horizontal demand curve over relevant output ranges at the price level established by the market (see Exhibit 5). The seller can offer any quantity at this set market price without affecting price. In contrast, **imperfect competition** is where an individual firm has enough share of the market (or can control a certain segment of the market) and is therefore able to exert some influence over price. Instead of a large number of competing firms, imperfect competition involves a smaller number of firms in the market relative to perfect competition

and in the extreme case only one firm (i.e., monopoly). Under any form of imperfect competition, the individual seller confronts a negatively sloped demand curve, where price and the quantity demanded by consumers are inversely related. In this case, price to the firm declines when a greater quantity is offered to the market; price to the firm increases when a lower quantity is offered to the market. This is shown in Exhibits 6 and 7.

Exhibit 4 Total, Average, and Marginal Revenue under Perfect Competition

Quantity Sold (Q)	Price (P)	Total Revenue (TR)	Average Revenue (AR)	Marginal Revenue (MR)
0	100	0	—	—
1	100	100	100	100
2	100	200	100	100
3	100	300	100	100
4	100	400	100	100
5	100	500	100	100
6	100	600	100	100
7	100	700	100	100
8	100	800	100	100
9	100	900	100	100
10	100	1,000	100	100

The **quantity** or **quantity demanded** variable is the amount of the product that consumers are willing and able to buy at each price level. The quantity sold can be affected by the business through such activities as sales promotion, advertising, and competitive positioning of the product that would take place under the market model of imperfect competition. Under perfect competition, however, total quantity in the market is influenced strictly by price, while non-price factors are not important. Once consumer preferences are established in the market, price determines the quantity demanded by buyers. Together, price and quantity constitute the firm's demand curve, which becomes the basis for calculating the total, average, and marginal revenue.

In Exhibit 4, **price** is the market price as established by the interactions of the market demand and supply factors. Since the firm is a price taker, price is fixed at 100 at all levels of output.

Total revenue (TR) is tabulated as price times the quantity of units sold. At 1 unit TR is 100 (calculated as 100×1 unit); at 10 units it is 1,000 (calculated as 100×10 units). At zero quantity, obviously, total revenue is always zero. Under perfect competition, for each increment in quantity, total revenue increases by the price level, which is constant to the firm. This relationship is shown in Exhibit 4—where the increase in total revenue from one quantity to the next equals 100, which is equal to the price.

Average revenue (AR) is quantity sold divided into total revenue. The mathematical outcome of this calculation is simply the price that the firm receives in the market for selling a given quantity. For any firm that sells at a uniform price, average revenue will equal price. For example, AR at 3 units is 100 (calculated as $300 \div 3$ units); at 8 units it is also 100 (calculated as $800 \div 8$ units).

Marginal revenue (MR) is the change in total revenue divided by the change in quantity sold; it is simply the additional revenue from selling one more unit. For example, in Exhibit 4, MR at 4 units is 100 [calculated as $(400 - 300) \div (4 - 3)$]; at

9 units it is also 100 [calculated as $(900 - 800) \div (9 - 8)$]. In a competitive market in which price is constant to the individual firm regardless of the amount of output offered, marginal revenue is equal to average revenue, where both are the same as the market price. Reviewing the revenue data in Exhibit 4, price, average revenue, and marginal revenue are all equal to 100. In the case of imperfect competition, MR declines with greater output and is less than AR at any positive quantity level, as will become clear with Exhibit 7.

Exhibit 5 graphically displays the revenue data from Exhibit 4. For an individual firm operating in a market setting of perfect competition, MR equals AR and both are equal to a price that stays the same across all levels of output. Because price is fixed to the individual seller, the firm's demand curve is a horizontal line at the point where the market sets the price. In Exhibit 5, at a price of 100, $P_1 = MR_1 = AR_1 = Demand_1$. Marginal revenue, average revenue, and the firm's price remain constant until market demand and supply factors cause a change in price. For instance, if price increases to 200 because of an increase in market demand, the firm's demand curve shifts from $Demand_1$ to $Demand_2$ with corresponding increases in MR and AR as well. Total revenue increases from TR_1 to TR_2 when price increases from 100 to 200. At a price of 100, total revenue at 10 units is 1,000; however, at a price of 200, total revenue would be 2,000 for 10 units.

Exhibit 5 Total Revenue, Average Revenue, and Marginal Revenue under Perfect Competition

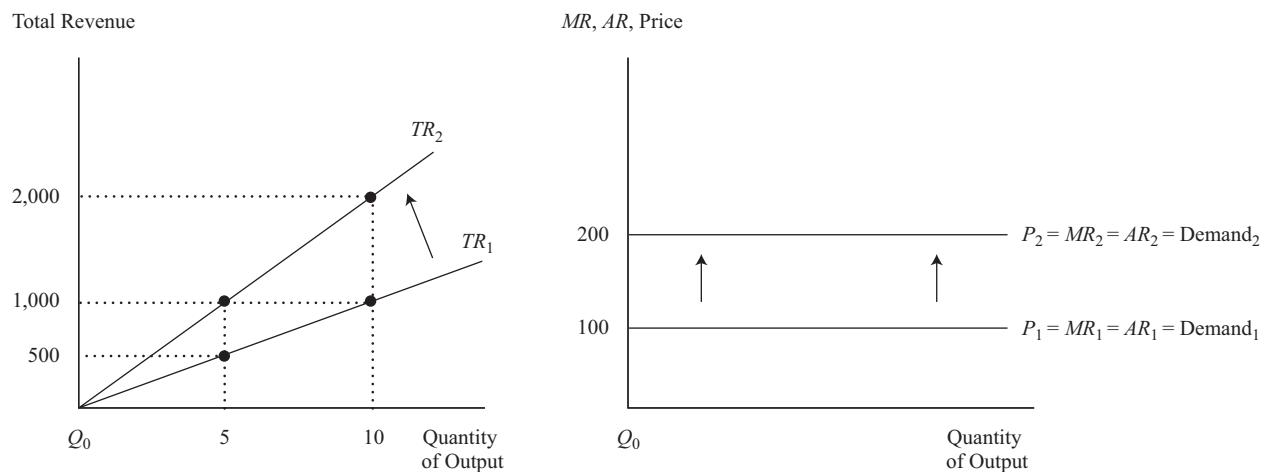
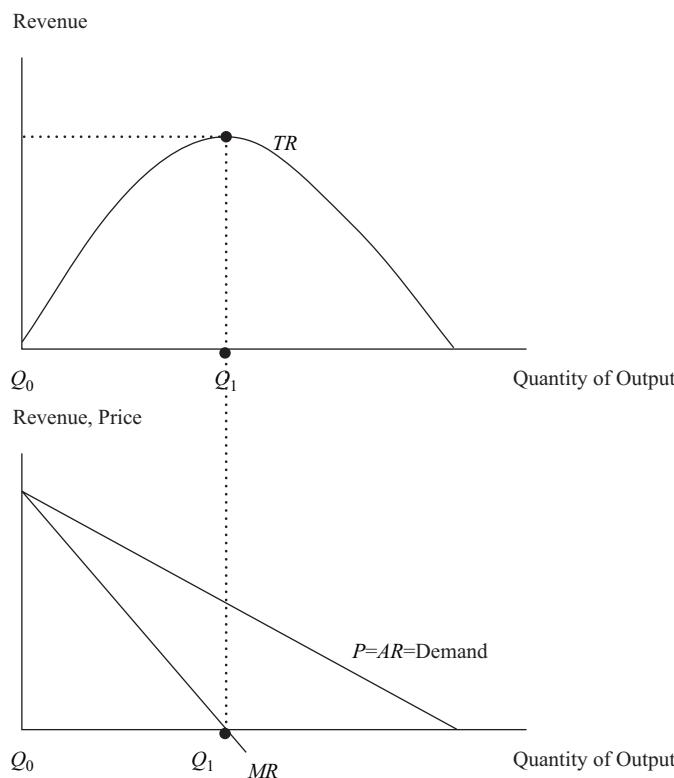


Exhibit 6 graphically illustrates the general shapes and relationships for TR , AR , and MR under imperfect competition. MR is positioned below the price and AR lines. TR peaks when MR equals zero at point Q_1 .

Exhibit 6 Total Revenue, Average Revenue, and Marginal Revenue under Imperfect Competition

EXAMPLE 2
Calculation and Interpretation of Total, Average, and Marginal Revenue under Imperfect Competition

Given quantity and price data in the first two columns of Exhibit 7, total revenue, average revenue, and marginal revenue can be calculated for a firm that operates under imperfect competition.

Exhibit 7

Quantity (Q)	Price (P)	Total Revenue (TR)	Average Revenue (AR)	Marginal Revenue (MR)
0	100	0	—	—
1	99	99	99	99
2	98	196	98	97
3	97	291	97	95
4	96	384	96	93
5	95	475	95	91
6	94	564	94	89
7	93	651	93	87
8	92	736	92	85

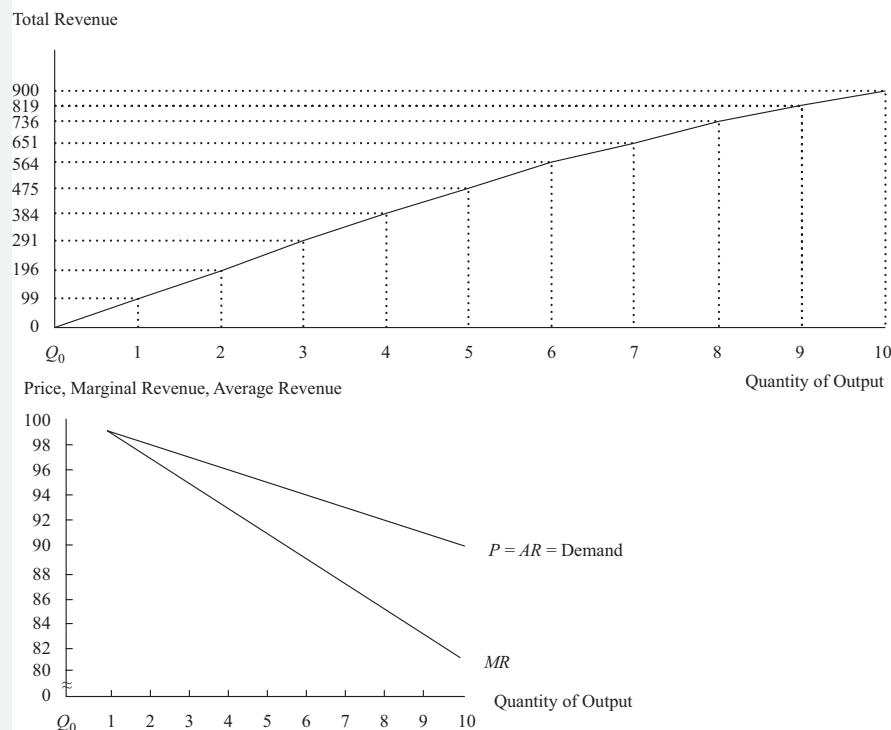
Exhibit 7 (Continued)

Quantity (Q)	Price (P)	Total Revenue (TR)	Average Revenue (AR)	Marginal Revenue (MR)
9	91	819	91	83
10	90	900	90	81

Describe how total revenue, average revenue, and marginal revenue change as quantity sold increases from 0 to 10 units.

Solution:

Total revenue increases with a greater quantity, but the rate of increase in *TR* (as measured by marginal revenue) declines as quantity increases. Average revenue and marginal revenue decrease when output increases, with *MR* falling faster than price and *AR*. Average revenue is equal to price at each quantity level. Exhibit 8 shows the relationships among the revenue variables presented in Exhibit 7.

Exhibit 8 Total Revenue, Average Revenue, and Marginal Revenue for Exhibit 7 Data

3.1.2 Factors of Production

Revenue generation occurs when output is sold in the market. However, costs are incurred before revenue generation takes place as the firm purchases resources, or what are commonly known as the factors of production, in order to produce a product or service that will be offered for sale to consumers. Factors of production, the inputs to the production of goods and services, include:

- *land*, as in the site location of the business;
- *labor*, which consists of the inputs of skilled and unskilled workers as well as the inputs of firms' managers;
- *capital*, which in this context refers to *physical capital*—such tangible goods as equipment, tools, and buildings. Capital goods are distinguished as inputs to production that are themselves produced goods; and
- *materials*, which in this context refers to any goods the business buys as inputs to its production process.¹

For example, a business that produces solid wood office desks needs to acquire lumber and hardware accessories as raw materials and hire workers to construct and assemble the desks using power tools and equipment. The factors of production are the inputs to the firm's process of producing and selling a product or service where the goal of the firm is to maximize profit by satisfying the demand of consumers. The types and quantities of resources or factors used in production, their respective prices, and how efficiently they are employed in the production process determine the cost component of the profit equation.

Clearly, in order to produce output, the firm needs to employ factors of production. While firms may use many different types of labor, capital, raw materials, and land, an analyst may find it more convenient to limit attention to a more simplified process in which only the two factors, capital and labor, are employed. The relationship between the flow of output and the two factors of production is called the **production function**, and it is represented generally as:

$$Q = f(K, L) \quad (5)$$

where Q is the quantity of output, K is capital, and L is labor. The inputs are subject to the constraint that $K \geq 0$ and $L \geq 0$. A more general production function is stated as:

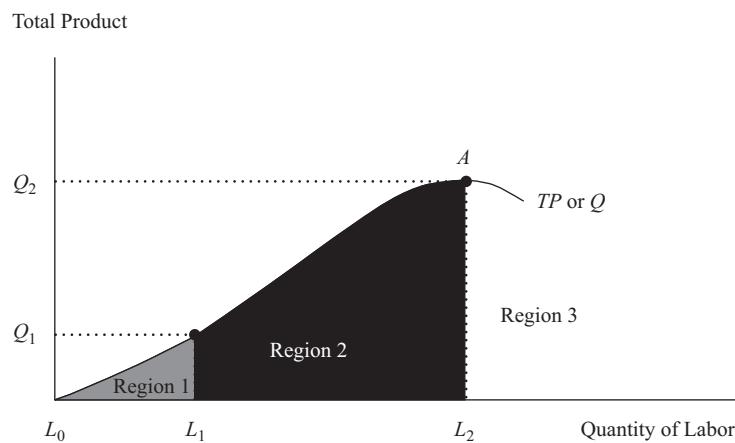
$$Q = f(x_1, x_2, \dots, x_n) \quad (6)$$

where x_i represents the quantity of the i th input subject to $x_i \geq 0$ for n number of different inputs. Exhibit 9 illustrates the shape of a typical input–output relationship using labor (L) as the only variable input (all other input factors are held constant). The production function has three distinct regions where both the direction of change and the rate of change in **total product** (TP or Q , quantity of output) vary as production changes. Regions 1 and 2 have positive changes in TP as labor is added, but the change turns negative in Region 3. Moreover, in Region 1 ($L_0 - L_1$), TP is increasing at an increasing rate, typically because specialization allows laborers to become increasingly productive. In Region 2, however, ($L_1 - L_2$), TP is increasing at a decreasing rate because capital is fixed, and labor experiences diminishing marginal returns. The firm would want to avoid Region 3 if at all possible because total product

¹ Because this factor may include such processed materials as steel and plastic that the firm purchases as inputs to production, the name *materials* was chosen in preference to another, traditional, name for this factor, *raw materials*. Candidates may encounter a number of variations in the classification and terminology for the factors of production.

or quantity would be declining rather than increasing with additional input: There is so little capital per unit of labor that additional laborers would possibly “get in each other’s way”. Point A is where TP is maximized.

Exhibit 9 A Firm’s Production Function



EXAMPLE 3

Factors of Production

A group of business investors are in the process of forming a new enterprise that will manufacture shipping containers to be used in international trade.

- 1 What decisions about factors of production must the start-up firm make in beginning operations?
- 2 What objective should guide the firm in its purchase and use of the production factors?

Solution to 1:

The entrepreneurs must decide where to locate the manufacturing facility in terms of an accessible site (land) and building (physical capital), what to use in the construction of the containers (materials), and what labor input to use.

Solution to 2:

Overall, any decision involving the input factors should focus on how that decision affects costs, profitability, and risk—such that shareholders’ wealth is maximized.

3.1.3 Total, Average, Marginal, Fixed, and Variable Costs

Exhibit 10 shows the graphical relationships among total costs, total fixed cost, and total variable cost. The curve for total costs is a parallel shift of the total variable cost curve and always lies above the total variable cost curve by the amount of total fixed cost. At zero production, total costs are equal to total fixed cost because total variable cost at this output level is zero.

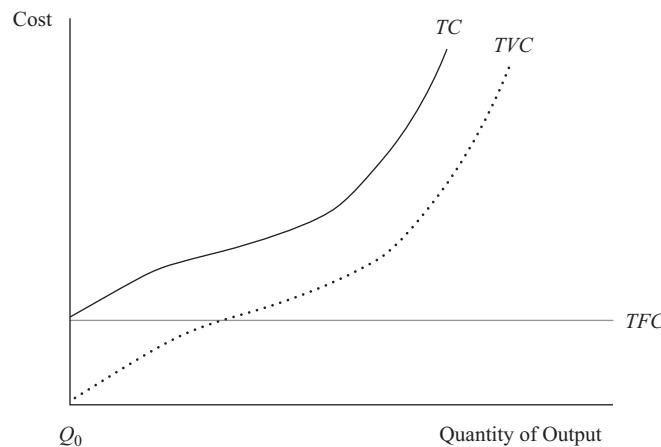
Exhibit 10 Total Costs, Total Variable Cost, and Total Fixed Cost

Exhibit 11 shows the cost curve relationships among ATC , AVC , and AFC in the short run. (In the long run, the firm will have different ATC , AVC , and AFC cost curves when all inputs are variable, including technology, plant size, and physical capital.) The difference between ATC and AVC at any output quantity is the amount of AFC . For example, at Q_1 the distance between ATC and AVC is measured by the value of A , which equals the amount of fixed cost as measured by amount B at Q_1 . Similarly, at Q_2 , the distance between ATC and AVC of X equals amount Y of AFC . The vertical distance between ATC and AVC is exactly equal to the height of AFC at each quantity. Both average total cost and average variable cost take on a bowl-shaped pattern in which each curve initially declines, reaches a minimum-cost output level, and then increases after that point. Point S , which corresponds to Q_{AVC} , is the minimum point on the AVC (such as 2 units in Exhibit 13). Similarly, point T , which corresponds to Q_{ATC} , is the minimum point on ATC (such as 3 units in Exhibit 13). As shown in Exhibit 11, when output increases, average fixed cost declines as AFC approaches the horizontal quantity axis.

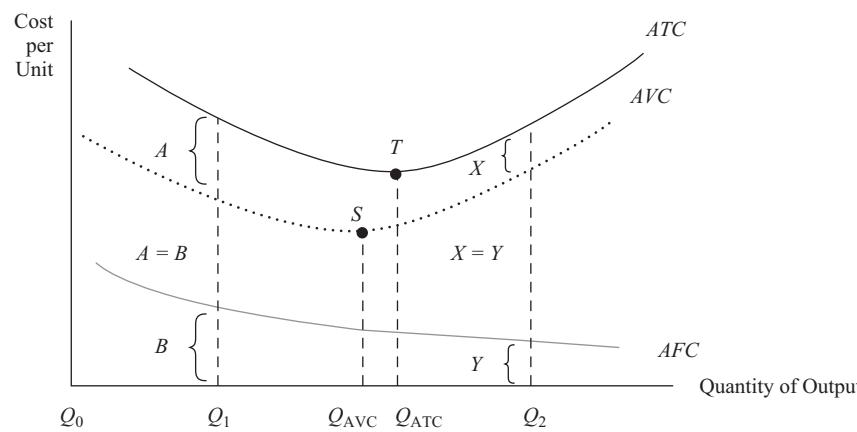
Exhibit 11 Average Total Cost, Average Variable Cost, and Average Fixed Cost

Exhibit 12 displays the cost curve relationships for ATC , AVC , and MC in the short run. The marginal cost curve intersects both the ATC and AVC at their respective minimum points. This occurs at points S and T , which correspond to Q_{AVC} and Q_{ATC} , respectively. Mathematically, when marginal cost is less than average variable cost, AVC will be decreasing. The opposite occurs when MC is greater than AVC . For example, in Exhibit 13, AVC begins to increase beyond 2 units, where MC exceeds AVC . The same relationship holds true for MC and ATC . Referring again to Exhibit 13, ATC declines up to 3 units when MC is less than ATC . After 3 units, ATC increases as MC exceeds ATC . Initially, the marginal cost curve declines, but at some point it begins to increase in reflection of an increasing rate of change in total costs as the firm produces more output. Point R (Exhibit 12), which corresponds to Q_{MC} , is the minimum point on the marginal cost curve.

Exhibit 12 Average Total Cost, Average Variable Cost, and Marginal Cost

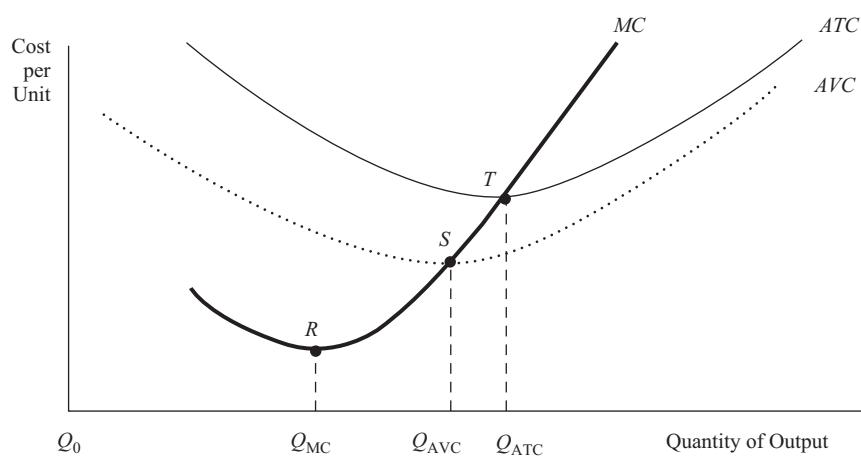


Exhibit 13 shows an example of how total, average, and marginal costs are derived. Total costs are calculated by summing total fixed cost and total variable cost. Marginal cost is derived by taking the change in total costs as the quantity variable changes.

Exhibit 13 Total, Average, Marginal, Fixed, and Variable Costs

Quantity (Q)	Total Fixed Cost ^a (TFC)	Average Fixed Cost (AFC)	Total Variable Cost (TVC)	Average Variable Cost (AVC)	Total Costs (TC)	Average Total Cost (ATC)	Marginal Cost (MC)
0	100	—	0	—	100	—	—
1	100	100.0	50	50.0	150	150.0	50
2	100	50.0	75	37.5	175	87.5	25
3	100	33.3	125	41.7	225	75.0	50
4	100	25.0	210	52.5	310	77.5	85
5	100	20.0	300	60.0	400	80.0	90
6	100	16.7	450	75.0	550	91.7	150
7	100	14.3	650	92.9	750	107.1	200
8	100	12.5	900	112.5	1,000	125.0	250

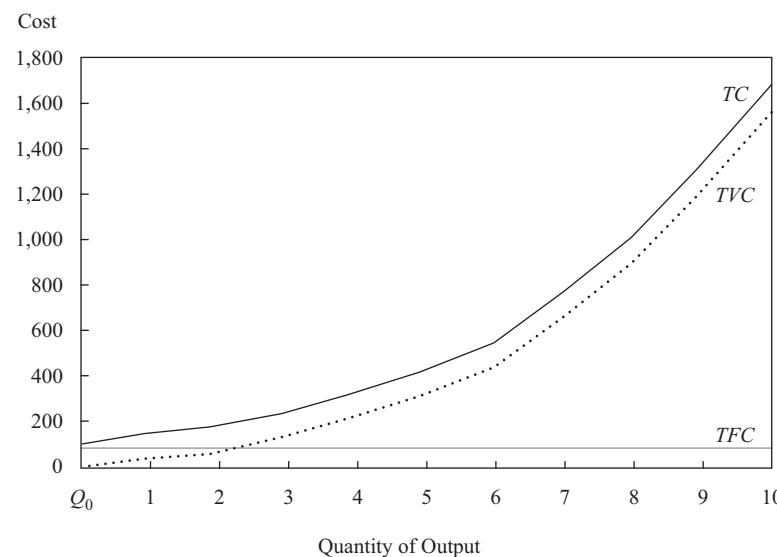
(continued)

Exhibit 13 (Continued)

Quantity (Q)	Total Fixed Cost ^a (TFC)	Average Fixed Cost (AFC)	Total Variable Cost (TVC)	Average Variable Cost (AVC)	Total Costs (TC)	Average Total Cost (ATC)	Marginal Cost (MC)
9	100	11.1	1,200	133.3	1,300	144.4	300
10	100	10.0	1,550	155.0	1,650	165.0	350

^a Includes all opportunity costs.

Exhibit 14 graphically displays the data for total costs, total variable cost, and total fixed cost from the table in Exhibit 13.

Exhibit 14 Total Costs, Total Variable Cost, and Total Fixed Cost for Exhibit 13 Data

Total costs (TC) are the summation of all costs, where costs are classified according to fixed or variable. Total costs increase as the firm expands output and decrease when production is cut. The rate of increase in total costs declines up to a certain output level and, thereafter, accelerates as the firm gets closer to full utilization of capacity. The rate of change in total costs mirrors the rate of change in total variable cost. In Exhibit 13, TC at 5 units is 400—of which 300 is variable cost and 100 is fixed cost. At 10 units, total costs are 1,650, which is the sum of 1,550 in variable cost and 100 in fixed cost.

Total fixed cost (TFC) is the summation of all expenses that do not change when production varies. It can be a sunk or unavoidable cost that a firm has to cover whether it produces anything or not, or it can be a cost that stays the same over a range of production but can change to another constant level when production moves outside of that range. The latter is referred to as a **quasi-fixed cost**, although it remains categorized as part of TFC . Examples of fixed costs are debt service, real estate lease agreements, and rental contracts. Quasi-fixed cost examples would be certain utilities and administrative salaries that could be avoided or be lower when output is zero but would assume higher constant values over different production ranges. Normal

profit is considered to be a fixed cost because it is a return required by investors on their equity capital regardless of output level. At zero output, total costs are always equal to the amount of total fixed cost that is incurred at this production point. In Exhibit 13, total fixed cost remains at 100 throughout the entire production range.

Other fixed costs evolve primarily from investments in such fixed assets as real estate, production facilities, and equipment. As a firm grows in size, fixed asset expansion occurs along with a related increase in fixed cost. However, fixed cost cannot be arbitrarily cut when production declines. Regardless of the volume of output, an investment in a given level of fixed assets locks the firm into a certain amount of fixed cost that is used to finance the physical capital base, technology, and other capital assets. When a firm downsizes, the last expense to be cut is usually fixed cost.

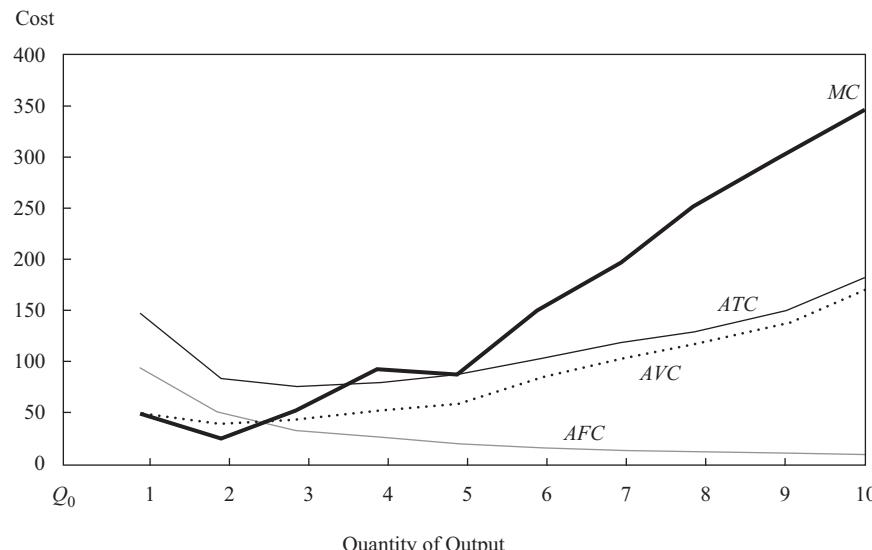
Total variable cost (TVC), which is the summation of all variable expenses, has a direct relationship with quantity. When quantity increases, total variable cost increases; total variable cost declines when quantity decreases. At zero production, total variable cost is always zero. Variable cost examples are payments for labor, raw materials, and supplies. As indicated above, total costs mirror total variable cost, with the difference being a constant fixed cost. The change in total variable cost (which defines marginal cost) declines up to a certain output point and then increases as production approaches capacity limits. In Exhibit 13, total variable cost increases with an increase in quantity. However, the change from 1 to 2 units is 25, calculated as $(75 - 50)$; the change from 9 to 10 units is 350, calculated as $(1,550 - 1,200)$.

Another approach to calculating total variable cost is to determine the variable cost per unit of output and multiply this cost figure by the number of production units. Per unit variable cost is the cost of producing each unit exclusive of any fixed cost allocation to production units. One can assign variable cost individually to units or derive an average variable cost per unit.

Whenever a firm initiates a downsizing, retrenchment, or defensive strategy, variable cost is the first to be considered for reduction given its variability with output. However, variable cost is reducible only so far because all firms have to maintain a minimum amount of labor and other variable resources to function effectively.

Exhibit 15 illustrates the relationships among marginal cost, average total cost, average variable cost, and average fixed cost for the data presented in Exhibit 13.

Exhibit 15 Average Total Cost, Average Variable Cost, Average Fixed Cost, and Marginal Cost for Exhibit 13 Data



Dividing total fixed cost by quantity yields **average fixed cost** (*AFC*), which decreases throughout the production span. A declining average fixed cost reflects spreading a constant cost over more and more production units. At high production volumes, *AFC* may be so low that it is a small proportion of average total cost. In Exhibit 13, *AFC* declines from 100 at 1 unit, to 20 at 5 units, and then to 10 at an output level of 10 units.

Average variable cost (*AVC*) is derived by dividing total variable cost by quantity. For example, average variable cost at 5 units is $(300 \div 5)$ or 60. Over an initial range of production, average variable cost declines and then reaches a minimum point. Thereafter, cost increases as the firm utilizes more of its production capacity. This higher cost results primarily from production constraints imposed by the fixed assets at higher volume levels. The minimum point on the *AVC* coincides with the lowest average variable cost. However, the minimum point on the *AVC* does not correspond to the least-cost quantity for average total cost. In Exhibit 13, average variable cost is minimized at 2 units, whereas average total cost is the lowest at 3 units.

Average total cost (*ATC*) is calculated by dividing total costs by quantity or by summing average fixed cost and average variable cost. For instance, in Exhibit 13, at 8 units *ATC* is 125 [calculated as $(1,000 \div 8)$ or $(AFC + AVC = 12.5 + 112.5)$]. Average total cost is often referenced as per-unit cost and is frequently called average cost. The minimum point on the average total cost curve defines the output level that has the least cost. The cost-minimizing behavior of the firm would dictate operating at the minimum point on its *ATC* curve. However, the quantity that maximizes profit (such as Q_3 in Exhibit 17) may not correspond to the *ATC*-minimum point. The minimum point on the *ATC* curve is consistent with maximizing profit per-unit, but it is not necessarily consistent with maximizing total profit. In Exhibit 13, the least-cost point of production is 3 units; *ATC* is 75, derived as $[(225 \div 3) \text{ or } (33.3 + 41.7)]$. Any other production level results in a higher *ATC*.

Marginal cost (*MC*) is the change in total cost divided by the change in quantity. Marginal cost also can be calculated by taking the change in total variable cost and dividing by the change in quantity. It represents the cost of producing an additional unit. For example, at 9 units marginal cost is 300, calculated as $[(1,300 - 1,000) \div (9 - 8)]$. Marginal cost follows a J-shaped pattern whereby cost initially declines but turns higher at some point in reflection of rising costs at higher production volumes. In Exhibit 13, *MC* is the lowest at 2 units of output with a value of 25, derived as $[(175 - 150) \div (2 - 1)]$.

EXAMPLE 4

Calculation and Interpretation of Total, Average, Marginal, Fixed, and Variable Costs

The first three columns of Exhibit 16 display data on quantity, total fixed cost, and total variable cost, which are used to calculate total costs, average fixed cost, average variable cost, average total cost, and marginal cost. Interpret the results for total, average, marginal, fixed, and variable costs.

Exhibit 16

<i>Q</i>	<i>TFC^a</i>	<i>TVC</i>	<i>AFC</i>	<i>AVC</i>	<i>TC</i>	<i>ATC</i>	<i>MC</i>
0	5,000	0	—	—	5,000	—	—
1	5,000	2,000	5,000.0	2,000	7,000	7,000.0	2,000
2	5,000	3,800	2,500.0	1,900	8,800	4,400.0	1,800

Exhibit 16 (Continued)

Q	TFC^a	TVC	AFC	AVC	TC	ATC	MC
3	5,000	5,400	1,666.7	1,800	10,400	3,466.7	1,600
4	5,000	8,000	1,250.0	2,000	13,000	3,250.0	2,600
5	5,000	11,000	1,000.0	2,200	16,000	3,200.0	3,000
6	5,000	15,000	833.3	2,500	20,000	3,333.3	4,000
7	5,000	21,000	714.3	3,000	26,000	3,714.3	6,000
8	5,000	28,800	625.0	3,600	33,800	4,225.0	7,800
9	5,000	38,700	555.6	4,300	43,700	4,855.6	9,900
10	5,000	51,000	500.0	5,100	56,000	5,600.0	12,300

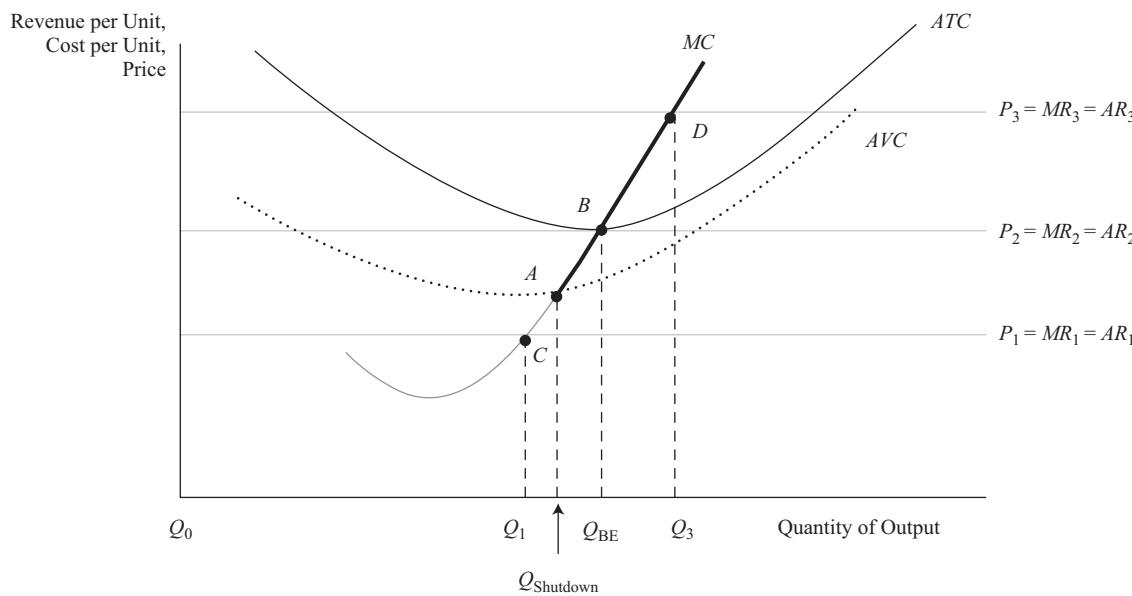
^a Includes all opportunity costs

Solution:

Total fixed cost remains unchanged at 5,000 throughout the entire production range, while average fixed cost continuously declines from 5,000 at one unit to 500 by 10 units. Both average variable cost and marginal cost initially decline and then reach their lowest level at 3 units, with costs of 1,800 and 1,600, respectively. Beyond 3 units, both average variable cost and marginal cost increase, indicating that the cost of production rises with greater output. The least-cost point for average total cost is 3,200 at 5 units. At zero output, total costs are 5,000, which equal the amount of total fixed cost.

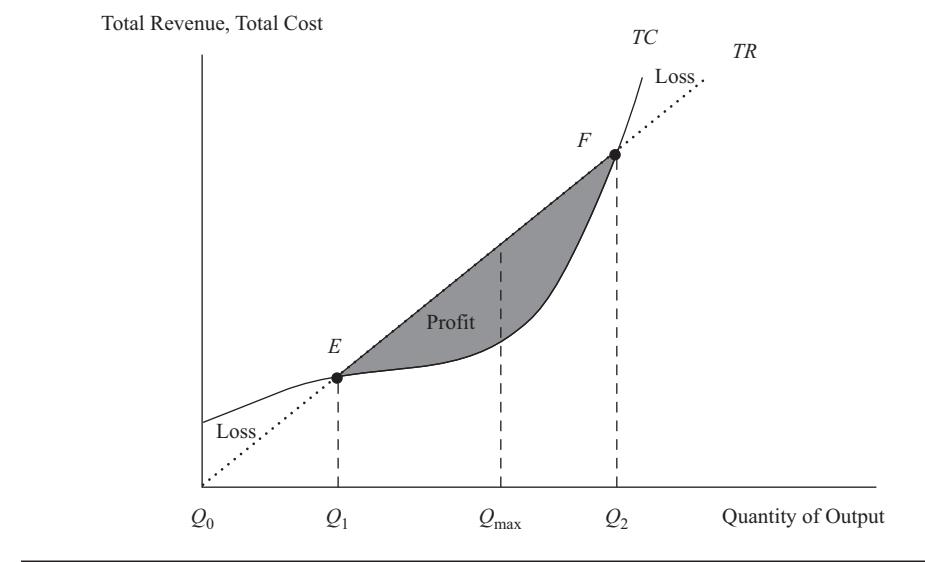
Exhibit 17 displays the firm's supply curve, shutdown point, and breakeven level of operation under perfect competition in the short run. The firm's **short-run supply curve** is the bold section of the marginal cost curve that lies above the minimum point (point A) on the average variable cost curve. If the firm operates below this point (for example between C and A), it shuts down because of its inability to cover variable costs in full. Between points A and B, the firm can operate in the short run because it is meeting variable cost payments even though it is unable to cover all of its fixed costs. In the long run, however, the firm is not able to survive if fixed costs are not completely covered. Any operating point above point B (the minimum point on ATC), such as point D, generates an economic profit.

A firm's **shutdown point** occurs when average revenue is less than average variable cost (any output below $Q_{shutdown}$), which corresponds to point A in Exhibit 17. Shutdown is defined as a situation in which the firm stops production but still confronts the payment of fixed costs in the short run as a business entity. In the short run, a business is capable of operating in a loss situation as long as it covers its variable costs even though it is not earning sufficient revenue to cover all fixed cost obligations. If variable costs cannot be covered in the short run ($P < AVC$), the firm will shut down operations and simply absorb the unavoidable fixed costs. This problem occurs at output Q_1 , which corresponds to point C where price is less than average variable cost. However, in the long run, to remain in business, the price must cover all costs. Therefore, in the long run, at any price below the breakeven point, the firm will exit the market, i.e., the firm will no longer participate in the market. Point D, which corresponds to output Q_3 , is a position where economic profit occurs because price is greater than ATC.

Exhibit 17 A Firm's Short-Run Supply Curve, Breakeven Point, and Shutdown Point under Perfect Competition


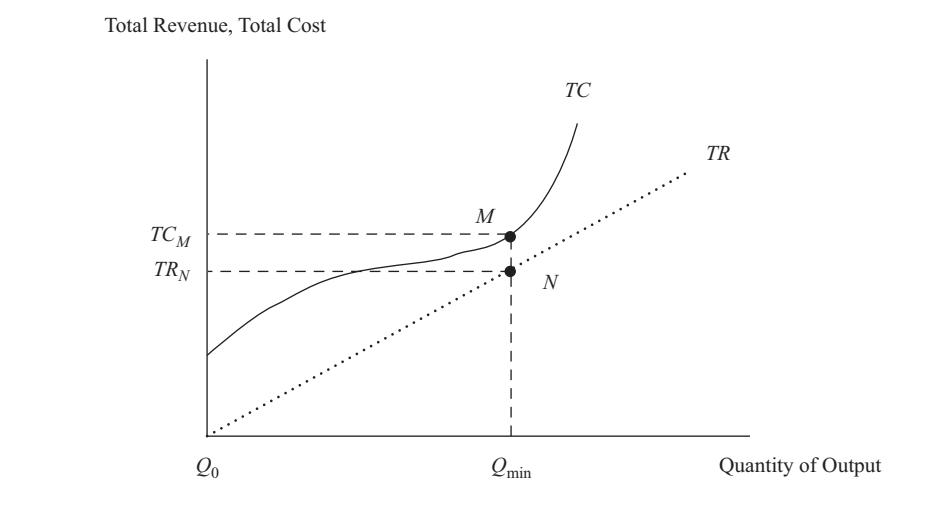
In the case of perfect competition, the **breakeven point** is the quantity where price, average revenue, and marginal revenue equal average total cost. It is also defined as the quantity where total revenue equals total costs. Firms strive to reach initial breakeven as soon as possible to avoid start-up losses for any extended period of time. When businesses are first established, there is an initial period where losses occur at low quantity levels. In Exhibit 17, the breakeven quantity occurs at output Q_{BE} , which corresponds to point B where price is tangent to the minimum point on the ATC. (Keep in mind that normal profit as an implicit cost is included in ATC as a fixed cost.)

Exhibit 18 shows the breakeven point under perfect competition using the total revenue–total cost approach. Actually, there are two breakeven points—lower (point E) and upper (point F). Below point E, the firm is losing money (economic losses), and beyond that point is the region of profitability (shaded area) that extends to the upper breakeven point. Within this profit area, a specific quantity (Q_{max}) maximizes profit as the largest difference between TR and TC. Point F is where the firm leaves the profit region and incurs economic losses again. This second region of economic losses develops when the firm's production begins to reach the limits of physical capacity, resulting in diminished productivity and an acceleration of costs. Obviously, the firm would not produce beyond Q_{max} because it is the optimal production point that maximizes profit.

Exhibit 18 A Firm's Breakeven Points Using Total Revenue and Total Costs under Perfect Competition


Breakeven points, profit regions, and economic loss ranges are influenced by demand and supply conditions, which change frequently according to the market behavior of consumers and firms. A high initial breakeven point is riskier than a low point because it takes a larger volume and, usually, a longer time to reach. However, at higher output levels it yields more return in compensation for this greater risk.

In the case where TC exceeds TR , as shown in Exhibit 19, the firm will want to minimize the economic loss (as long as $TR > TVC$), which is defined as the smallest difference between TC and TR . This occurs at Q_{min} , where the economic loss is calculated as $(TC_M - TR_N)$ on the vertical axis.

Exhibit 19 Loss Minimization Using Total Revenue and Total Costs


EXAMPLE 5

Breakeven Analysis and Profit Maximization When the Firm Faces a Negatively Sloped Demand under Imperfect Competition

The following revenue and cost information for a future period is presented in Exhibit 20 for WR International, a newly formed corporation that engages in the manufacturing of low-cost, pre-fabricated dwelling units for urban housing markets in emerging economies. (Note that quantity increments are in blocks of 10 for a 250 change in price.) The firm has few competitors in a market setting of imperfect competition.

- 1 How many units must WR International sell to initially break even?
- 2 Where is the region of profitability?
- 3 At what point will the firm maximize profit? At what points are there economic losses?

Exhibit 20

Quantity (Q)	Price (P)	Total Revenue (TR)	Total Costs (TC) ^a	Profit
0	10,000	0	100,000	(100,000)
10	9,750	97,500	170,000	(72,500)
20	9,500	190,000	240,000	(50,000)
30	9,250	277,500	300,000	(22,500)
40	9,000	360,000	360,000	0
50	8,750	437,500	420,000	17,500
60	8,500	510,000	480,000	30,000
70	8,250	577,500	550,000	27,500
80	8,000	640,000	640,000	0
90	7,750	697,500	710,000	(12,500)
100	7,500	750,000	800,000	(50,000)

^a Includes all opportunity costs

Solution to 1:

WR International will initially break even at 40 units of production, where TR and TC equal 360,000.

Solution to 2:

The region of profitability will range from 40 to 80 units. Any production quantity of less than 40 units and any quantity greater than 80 will result in an economic loss.

Solution to 3:

Maximum profit of 30,000 will occur at 60 units. Lower profit will occur at any output level that is higher or lower than 60 units. From zero quantity to 40 units and for quantities beyond 80 units, economic losses occur.

Given the relationships among total revenue, total variable cost, and total fixed cost, Exhibit 21 summarizes the decisions to operate, shut down production, or exit the market in both the short run and long run. As previously discussed, the firm must cover variable cost before fixed cost. In the short run, if total revenue cannot cover total variable cost, the firm shuts down production to minimize loss, which would equal the amount of fixed cost. If total variable cost exceeds total revenue in the long run, the firm will exit the market as a business entity to avoid the loss associated with fixed cost at zero production. By terminating business operations through market exit, investors escape the erosion in their equity capital from economic losses. When total revenue is enough to cover total variable cost but not all of total fixed cost, the firm can survive in the short run but will be unable to maintain financial solvency in the long run.

Exhibit 21

Revenue–Cost Relationship	Short-Run Decision	Long-Term Decision
$TR \geq TC$	Stay in market	Stay in market
$TR > TVC$ but $TR < TFC + TVC$	Stay in market	Exit market
$TR < TVC$	Shut down production to zero	Exit market

EXAMPLE 6

Shutdown Analysis

For the most recent financial reporting period, a business domiciled in Ecuador (which recognizes the US dollar as an official currency) has revenue of \$2 million and total costs of \$2.5 million, which are or can be broken down into total fixed cost of \$1 million and total variable cost of \$1.5 million. The net loss on the firm's income statement is reported as \$500,000 (ignoring tax implications). In prior periods, the firm had reported profits on its operations.

- 1 What decision should the firm make regarding operations over the short term?
- 2 What decision should the firm make regarding operations over the long term?
- 3 Assume the same business scenario except that revenue is now \$1.3 million, which creates a net loss of \$1.2 million. What decision should the firm make regarding operations in this case?

Solution to 1:

In the short run, the firm is able to cover all of its total variable cost but only half of its \$1 million in total fixed cost. If the business ceases to operate, its loss is \$1 million, the amount of total fixed cost, whereas the net loss by operating is minimized at \$500,000. The firm should attempt to operate by negotiating special arrangements with creditors to buy time to return operations back to profitability.

Solution to 2:

If the revenue shortfall is expected to persist over time, the firm should cease operations, liquidate assets, and pay debts to the extent possible. Any residual for shareholders would decrease the longer the firm is allowed to operate unprofitably.

Solution to 3:

The firm would minimize loss at \$1 million of total fixed cost by shutting down compared with continuing to do business where the loss is \$1.2 million. Shareholders will save \$200,000 in equity value by pursuing this option. Unquestionably, the business will have a rather short life expectancy if this loss situation were to continue.

When evaluating profitability, particularly of start-up firms and businesses using turnaround strategies, analysts should consider highlighting breakeven and shutdown points in their financial research. Identifying the unit sales levels where the firm enters or leaves the production range for profitability and where the firm can no longer function as a viable business entity provides invaluable insight to investment decisions.

3.1.4 Output Optimization and Maximization of Profit

Profit maximization occurs when

- the difference between total revenue (TR) and total costs (TC) is the greatest;
- marginal revenue (MR) equals marginal cost (MC); and
- the revenue value of the output from the last unit of input employed equals the cost of employing that input unit (as later developed in Equation 12).

All three approaches derive the same profit-maximizing output level. In the first approach, a firm starts by forecasting unit sales, which becomes the basis for estimates of future revenue and production costs. By comparing predicted total revenue to predicted total costs for different output levels, the firm targets the quantity that yields the greatest profit. When using the marginal revenue–marginal cost approach, the firm compares the change in predicted total revenue (MR) with the change in predicted total costs (MC) by unit of output. If MR exceeds MC , total profit is increased by producing more units because each successive unit adds more to total revenue than it does to total costs. If MC is greater than MR , total profit is decreased when additional units are produced. The point of profit maximization occurs where MR equals MC . The third method compares the estimated cost of each unit of input to that input's contribution with projected total revenue. If the increase in projected total revenue coming from the input unit exceeds its cost, a contribution to total profit is evident. In turn, this justifies further employment of that input. On the other hand, if the increase in projected total revenue does not cover the input unit's cost, total profit is diminished. Profit maximization based on the employment of inputs occurs where the next input unit for each type of resource used no longer makes any contribution to total profit.

Combining revenue and cost data from Exhibits 4 and 13, Exhibit 22 demonstrates the derivation of the optimal output level that maximizes profit for a firm under perfect competition. Profit is calculated as the difference between total revenue and total costs. At zero production, an economic loss of 100 occurs, which is equivalent to total fixed cost. Upon initial production, the firm incurs an economic loss of 50 on the first unit but breaks even by unit 2. The region of profitability ranges from 2 to 6 units. Within this domain, total profit is maximized in the amount of 100 at 5 units of output. No other quantity level yields a higher profit. At this 5-unit level,

marginal revenue exceeds marginal cost. But at unit 6, marginal revenue is less than marginal cost, which results in a lower profit because unit 6 costs more to produce than what it generates in revenue. Unit 6 costs 150 to produce but contributes only 100 to total revenue, which yields a 50 loss on that unit. As a result, profit drops from 100 to 50. At unit 7 and beyond, the firm begins to lose money again as it passes the upper breakeven mark and enters a second economic loss zone.

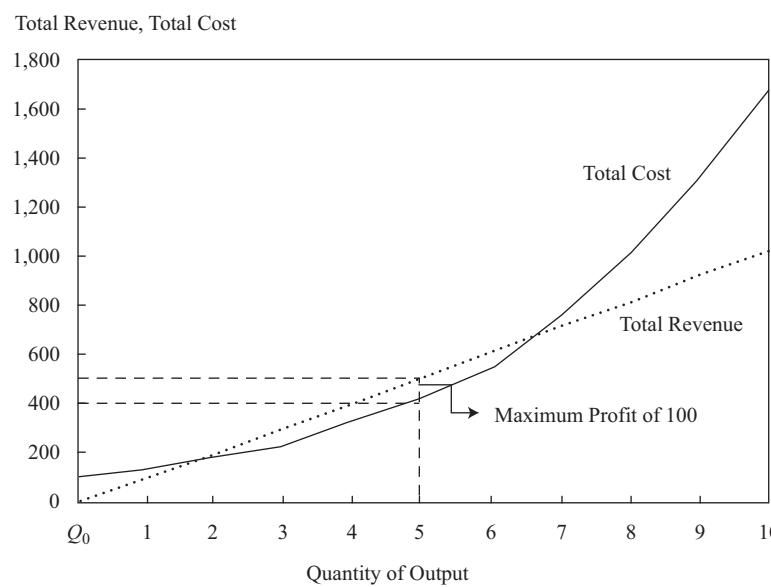
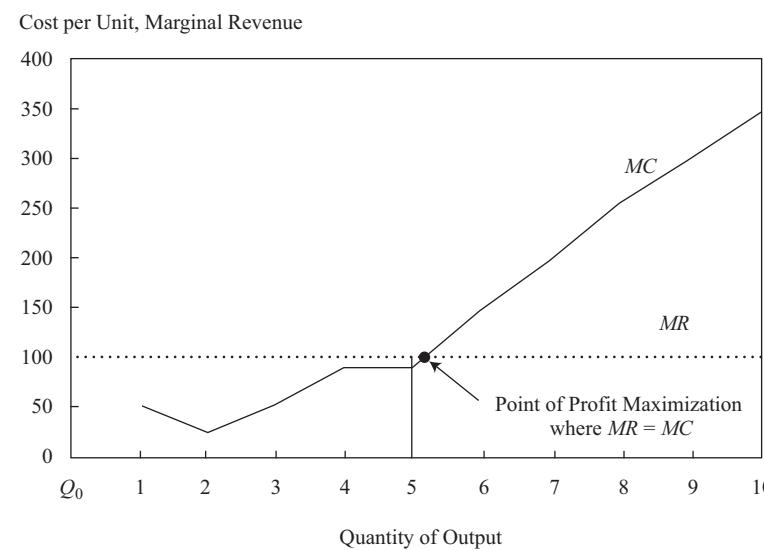
Exhibit 22 Profit Maximization under Perfect Competition

Quantity (Q)	Price (P)	Total Revenue (TR)	Total Costs (TC) ^a	Profit (P)	Marginal Revenue (MR)	Marginal Cost (MC)
0	100	0	100	(100)	—	—
1	100	100	150	(50)	100	50
2	100	200	175	25	100	25
3	100	300	225	75	100	50
4	100	400	310	90	100	85
5	100	500	400	100	100	90
6	100	600	550	50	100	150
7	100	700	750	(50)	100	200
8	100	800	1,000	(200)	100	250
9	100	900	1,300	(400)	100	300
10	100	1,000	1,650	(650)	100	350

^a Includes all opportunity costs

Exhibits 23 and 24 display the data from Exhibit 22 to illustrate profit maximization under perfect competition using the $(TR - TC)$ and $(MR = MC)$ approaches. Exhibit 24 highlights profit maximization based on comparing how much each unit of output costs to produce (MC) to how much each unit contributes to revenue (MR). Each unit up to and including unit 5 contributes to profit in that each unit's marginal revenue exceeds its marginal cost. Starting at unit 6 and thereafter, the marginal revenue for each unit is less than the marginal cost. This results in a reduction in profit. Profit maximization occurs where MR equals MC . In this case, the optimal decision for the firm using a comparison of MR and MC is to produce 5 units.²

² Marginal analysis is a common and valuable optimization tool that is used to determine the point of profit maximization and the optimal employment of resources. It is often said that the firm makes decisions “on the margin.”

Exhibit 23 Profit Maximization Using Total Revenue and Total Costs from Exhibit 22

Exhibit 24 Profit Maximization Using Marginal Revenue and Marginal Cost from Exhibit 22


It should be noted that under imperfect competition, the firm faces a negatively sloped demand curve. As the firm offers a greater quantity to the market, price decreases. In contrast, a firm under perfect competition has an insignificant share of the market and is able to sell more without impacting market price. Obviously, the type of market structure in which a firm operates as a seller has an impact on the firm's profit in terms of the price received when output levels vary.

EXAMPLE 7**Profit Maximization and the Breakeven Point under Imperfect Competition**

Exhibit 25 shows revenue and cost data for a firm that operates under the market structure of imperfect competition.

- 1 At what point does the firm break even over its production range in the short run?
- 2 What is the quantity that maximizes profit given total revenue and total costs?
- 3 Comparing marginal revenue and marginal cost, determine the quantity that maximizes profit.

Exhibit 25

<i>Q</i>	<i>P</i>	<i>TR</i>	<i>TC^a</i>	<i>Profit</i>	<i>MR</i>	<i>MC</i>
0	1,000	0	550	(550)	—	—
1	995	995	1,000	(5)	995	450
2	990	1,980	1,500	480	985	500
3	985	2,955	2,100	855	975	600
4	980	3,920	2,800	1,120	965	700
5	975	4,875	3,600	1,275	955	800
6	970	5,820	4,600	1,220	945	1,000
7	965	6,755	5,800	955	935	1,200
8	960	7,680	7,200	480	925	1,400
9	955	8,595	8,800	(205)	915	1,600
10	950	9,500	10,800	(1,300)	905	2,000

^a Includes all opportunity costs.

Solution to 1:

The breakeven point occurs between unit 1 and unit 2, where profit increases from (5) to 480.

Solution to 2:

At an output level of 5 units, the firm maximizes profit in the amount of 1,275, calculated as the difference between *TR* of 4,875 and *TC* of 3,600.

Solution to 3:

Profit maximization occurs at 5 units, where *MR* of 955 exceeds *MC* of 800, which yields a profit contribution of 155. However, at 6 units, *MR* of 945 is less than the *MC* of 1,000, resulting in a loss of 55 and a reduction in profit from 1,275 to 1,220.

Exhibit 26 summarizes the (*TR* – *TC*) and (*MR* = *MC*) profit-maximization approaches for firms operating under perfect competition. (Profit maximization using inputs is discussed in Section 3.2.2.)

Exhibit 26 Summary of Profit Maximization and Loss Minimization under Perfect Competition

Revenue–Cost Relationship	Actions by Firm
$TR = TC$ and $MR > MC$	Firm is operating at lower breakeven point; increase Q to enter profit territory.
$TR \geq TC$ and $MR = MC$	Firm is at maximum profit level; no change in Q .
$TR < TC$ and $TR \geq TVC$ but $(TR - TVC) < TFC$ (covering TVC but not TFC)	Find level of Q that minimizes loss in the short run; work toward finding a profitable Q in the long run; exit market if losses continue in the long run.
$TR < TVC$ (not covering TVC in full)	Shut down in the short run; exit market in the long run.
$TR = TC$ and $MR < MC$	Firm is operating at upper breakeven point; decrease Q to enter profit territory.

Profit acts as an efficient allocator of equity capital to investment opportunities whereby shareholders' wealth is increased. Equity flows from low-return business investments to high-return business investments as it seeks the greatest return potential on a risk-adjusted basis. Basic economic theory describes how consumer choice voiced through the price mechanism in competitive markets directs resources to their most efficient use according to what consumers need and want. In the end, it is profitability—which evolves from the interactions of demand and supply factors in product and resource markets—that decides where financial capital is employed.

3.1.5 Economies of Scale and Diseconomies of Scale

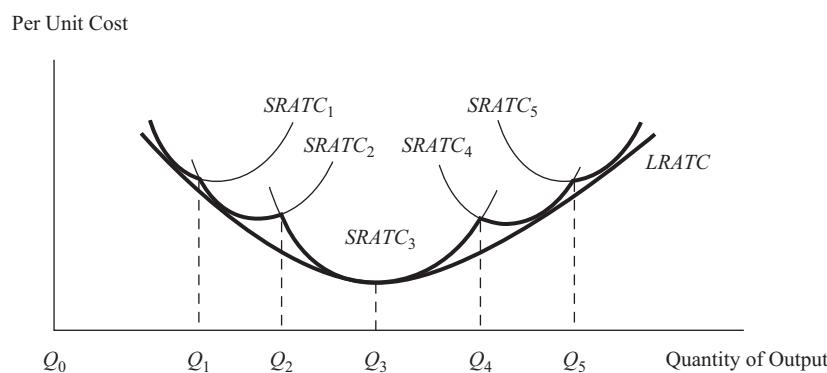
Rational behavior dictates that the firm select an operating size or scale that maximizes profit over any time frame. The time frame for the firm can be separated into the short run and long run based on the ability of the firm to adjust the quantities of the fixed resources it employs. The short run is defined as a time period in which at least one of the factors of production is fixed. The most likely inputs to be held constant in defining the short run are technology, physical capital, and plant size. Usually, a firm cannot change these inputs in a relatively short period of time given the inflexible nature of their use. The long run is defined as a time period in which all factors of production are variable, including technology, physical capital, and plant size. Additionally, in the long run, firms can enter or exit the market based on decisions regarding profitability. The long run is often referred to as the **planning horizon** in which the firm can choose the short-run position or optimal operating size that maximizes profit over time.

The time required for long-run adjustments varies by industry. For example, the long run for a small business using very little in the way of technology and physical capital may be less than a year. On the other hand, for a capital-intensive firm, the long run may be more than a decade. However, given enough time, all production factors are variable, which allows the firm to choose an operating size or plant capacity based on different technologies and physical capital. In this regard, costs and profits will differ between the short run and long run.

The fixed-input constraint in the short run along with input prices establish the firm's **short-run average total cost curve** (SRATC). This defines what the per-unit cost will be for any quantity in the short run. The SRATC and the demand for the firm's product determine short-run profit. The selection of technology, physical capital, and plant size is a key determinant of the short-run cost curve for the firm. As the firm switches to newer technologies and physical capital, a corresponding change in short-run costs occurs. In the long run, a firm has the opportunity for greater profit potential based on the ability to lower its costs through choices of more-efficient technology and physical capital and a wider selection of production capacities.

Exhibit 27 displays the **long-run average total cost curve** (LRATC), which is derived from the short-run average total cost curves that are available to the firm.³ The business has a choice of five technology-physical capital options and plant capacities over the long run, each with its own short-run cost curve. The LRATC consists of sections of these individual short-run cost curves. For example, from zero production to Q_1 output, $SRATC_1$ yields the lowest per unit cost. Between Q_1 and Q_2 , the lowest cost per unit is attainable with $SRATC_2$, which represents a larger production capacity. $SRATC_3$ and $SRATC_4$ would provide for the lowest average cost over time for output levels of $Q_2 - Q_4$ and $Q_4 - Q_5$, respectively. For any output greater than Q_5 , $SRATC_5$ becomes the preferred curve for minimizing average total cost. Tangentially connecting all of the least-cost SRATC segments by way of an envelope curve creates the LRATC. (Assuming an unlimited number of possible technologies, plant sizes, and physical capital combinations—and therefore a theoretically unlimited number of SRATC—the LRATC becomes a smooth curve rather than a segmented one as indicated by the bold segments of the five SRATC's in Exhibit 27.) The LRATC shows the lowest cost per unit at which output can be produced over a long period of time when the firm is able to make technology, plant size, and physical capital adjustments. If the same technologies and physical capital are available and adaptable to every firm in the industry, then all firms would have the same LRATC. However, firms could be at different positions on this homogenous LRATC depending on their operating size that is based on output.

Exhibit 27 Long-Run Average Total Cost Curve



Over the long run, as a business expands output, it can utilize more efficient technology and physical capital and take advantage of other factors to lower the costs of production. This development is referred to as **economies of scale** or **increasing returns to scale** as a firm moves to lower cost structures when it grows in size. Output increases by a larger proportion than the increase in inputs. The opposite effect can result after a certain volume level at which the business faces higher costs as it expands in size. This outcome is called **diseconomies of scale** or **decreasing returns to scale**, where the firm becomes less efficient with size. In this case, output increases by a smaller proportion than the increase in inputs. Diseconomies of scale often result from the firm becoming too large to be managed efficiently even though better technology can be increasing productivity within the business. Both economies and diseconomies

³ Some writers use short-run average cost (SRAC) and long-run average cost (LRAC) in the same sense as SRATC and LRATC, respectively.

of scale can occur at the same time; the impact on long-run average total cost depends on which dominates. If economies of scale dominate, LRATC decreases with increases in output; the reverse holds true when diseconomies of scale prevail.

Referring back to Exhibit 27, economies of scale occur from Q_0 (zero production) to output level Q_3 , where Q_3 is the cost-minimizing level of output for $SRATC_3$. It is evident throughout this production range that per-unit costs decline as the firm produces more. Over the production range of Q_3 to Q_5 , diseconomies of scale are occurring as per-unit costs increase when the firm expands output. Under perfect competition, given the five SRATC selections that are available to the firm throughout the production range $Q_0 - Q_5$, $SRATC_3$ is the optimal technology, plant capacity, and physical capital choice, with Q_3 being the target production size for the firm that would minimize cost over the long term.

Perfect competition forces the firm to operate at the minimum point on the LRATC because market price will be established at this level over the long run. If the firm is not operating at this least-cost point, its long-term viability will be threatened. The minimum point on the LRATC is referred to as the **minimum efficient scale** (MES). The MES is the optimal firm size under perfect competition over the long run where the firm can achieve cost competitiveness.

As the firm grows in size, economies of scale and a lower average total cost can result from the following factors:

- Division of labor and management in a large firm with numerous workers, where each worker can specialize in one task rather than perform many duties, as in the case of a small business (as such, workers in a large firm become more proficient at their jobs).
- Being able to afford more-expensive, yet more-efficient equipment and to adapt the latest in technology that increases productivity.
- Effectively reducing waste and lowering costs through marketable byproducts, less energy consumption, and enhanced quality control.
- Better use of market information and knowledge for more-effective managerial decision making.
- Discounted prices on resources when buying in larger quantities.

A classic example of a business that realizes economies of scale through greater physical capital investment is the electric utility. By expanding output capacity to accommodate a larger customer base, the utility company's per-unit cost will decline. Economies of scale help to explain why electric utilities have naturally evolved from localized entities to regional and multi-region enterprises. Wal-Mart, the world's largest retailer, is an example of a business that uses bulk purchasing power to obtain deep discounts from suppliers to keep costs and prices low. Wal-Mart also utilizes the latest in technology to monitor point-of-sale transactions to have timely market information to respond to changes in customer buying behavior. This leads to economies of scale through lower distribution and inventory costs.

The factors that can lead to diseconomies of scale, inefficiencies, and rising costs when a firm increases in size include:

- So large that it cannot be properly managed.
- Overlap and duplication of business functions and product lines.
- Higher resource prices because of supply constraints when buying inputs in large quantities.

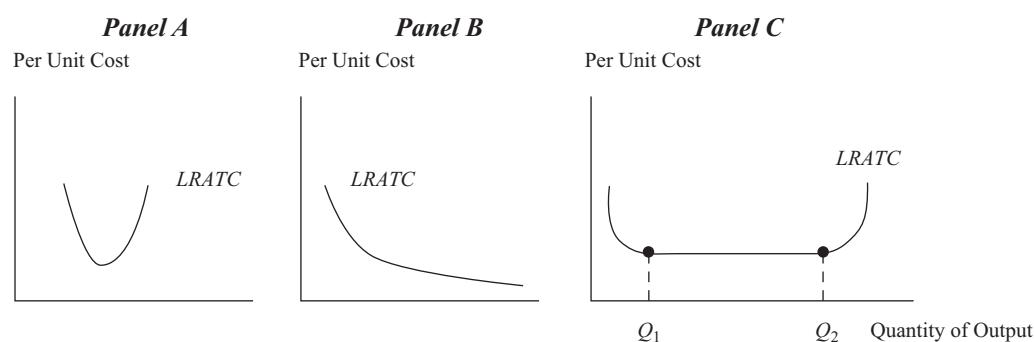
General Motors (GM) is an example of a business that has realized diseconomies of scale by becoming too large. Scale diseconomies have occurred through product overlap and duplication (i.e., similar or identical automobile models), where the fixed cost for these models is not spread over a large volume of output. (Recently, the company

has decided to discontinue various low-volume product models that overlapped with other models.) GM has numerous manufacturing plants throughout the world and sells vehicles in over a hundred countries. Given this geographical dispersion in production and sales, the company has had communication and management coordination problems, which have resulted in higher costs. Also, GM has had significantly higher labor costs when compared with its competitors. By being the largest producer in the market, it has been a target of labor unions for higher compensation and benefits packages relative to other firms.

Strategically, when a firm is operating in the economies of scale region, expanding production capacity will increase the firm's competitiveness through lower costs. Firm expansion is often facilitated with a growth or business combination (i.e., merger or acquisition) strategy. On the other hand, when a business is producing in the area of diseconomies of scale, the objective is to downsize to reduce costs and become more competitive. From an investment perspective, a firm operating at the minimum point of the industry LRATC under perfect competition should be valued higher than a firm that is not producing at this least-cost quantity.

The LRATC can take various forms given the development of new technology and growth prospects for an industry over the long term. Exhibit 28 displays examples of different average total cost curves that firms can realize over the long run. Panel A shows that scale economies dissipate rapidly at low output levels. This implies that a firm with a low volume of output can be more cost competitive than a firm that is producing a high output volume. Panel B indicates a lower and lower average cost over time as firm size increases. The larger the business, the more competitive it is and the greater its potential investment value. Finally, Panel C shows the case of **constant returns to scale** (i.e., output increases by the same proportion as the increase in inputs) over the range of production from Q_1 to Q_2 , indicating that size does not give a firm a competitive edge over another firm within this range. In other words, a firm that is producing the smaller output Q_1 has the same long-run average total cost as a firm producing the higher output Q_2 .

Exhibit 28 Types of Long-Run Average Total Cost Curves



EXAMPLE 8

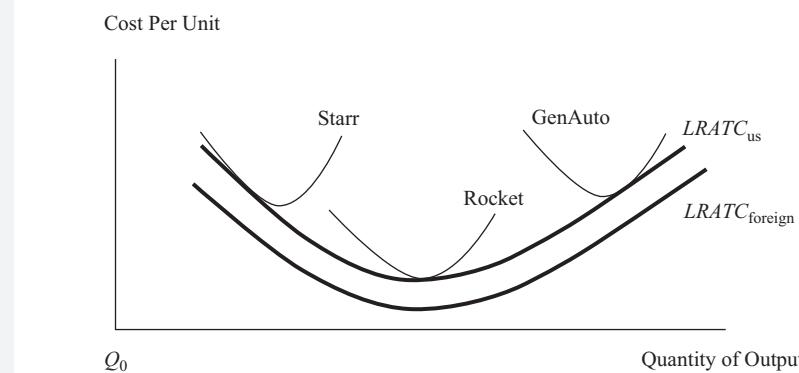
Long-Run Average Total Cost Curve

Exhibit 29 displays the long-run average total cost curve ($LRATC_{US}$) and the short-run average total cost curves for three hypothetical US-based automobile manufacturers—Starr Vehicles (Starr), Rocket Sports Cars (Rocket), and General Auto (GenAuto). The long-run average total cost curve ($LRATC_{foreign}$)

for foreign-owned automobile companies that compete in the US auto market is also indicated in the graph. (The market structure implicit in the exhibit is imperfect competition.)

To what extent are the cost relationships depicted in Exhibit 29 useful for an economic and financial analysis of the three US-based auto firms?

Exhibit 29



Solution:

First, it is observable that the foreign auto companies have a lower LRATC compared with that of the US automobile manufacturers. This competitive position places the US firms at a cost and possible pricing disadvantage in the market with the potential to lose market share to the lower-cost foreign competitors. Second, only Rocket operates at the minimum point of the $LRATC_{US}$, whereas GenAuto is situated in the region of diseconomies of scale and Starr is positioned in the economies of scale portion of the curve. To become more efficient and competitive, GenAuto needs to downsize and restructure, which means moving down the $LRATC_{US}$ curve to a smaller, yet lower-cost production volume. On the other hand, Starr has to grow in size to become more efficient and competitive by lowering per-unit costs.

From a long-term investment prospective and given its cost advantage, Rocket has the potential to create more investment value relative to GenAuto and Starr. Over the long run, if GenAuto and Starr can lower their average total costs, they will become more attractive to investors. On the other hand, if any or all of the three US auto companies cannot match the cost competitiveness of the foreign firms, they may be driven from the market. In the long run, the lower-cost foreign automakers pose a severe competitive challenge to the survival of the US manufacturers and their ability to maintain and grow shareholders' wealth.

3.1.6 Profit Maximization in the Short Run and Long Run

No matter the time span, the firm's supply behavior centers on the objective of profit maximization. In the short term, when technology and physical capital are fixed, maximum profit (or minimal loss) is determined where marginal cost equals marginal revenue (points A and B in Exhibit 30). Cases of profit maximization and loss minimization are illustrated in Exhibit 30 for a firm operating under perfect competition in the short run. In Panel A, the firm realizes economic profit because TR is greater than TC and price exceeds SRATC in the production range of Q_1-Q_2 . Q_{max} is the output level that maximizes economic profit. Panel B shows the case of loss minimization

because TC exceeds TR and $SRATC$ is above the price level. Q_{min} yields the least loss of all possible production quantities. Note that in this case, the short-run loss is still less than fixed cost, so the firm should continue operating in the short run.

Exhibit 30 Profit Maximization and Loss Minimization in the Short Run under Perfect Competition

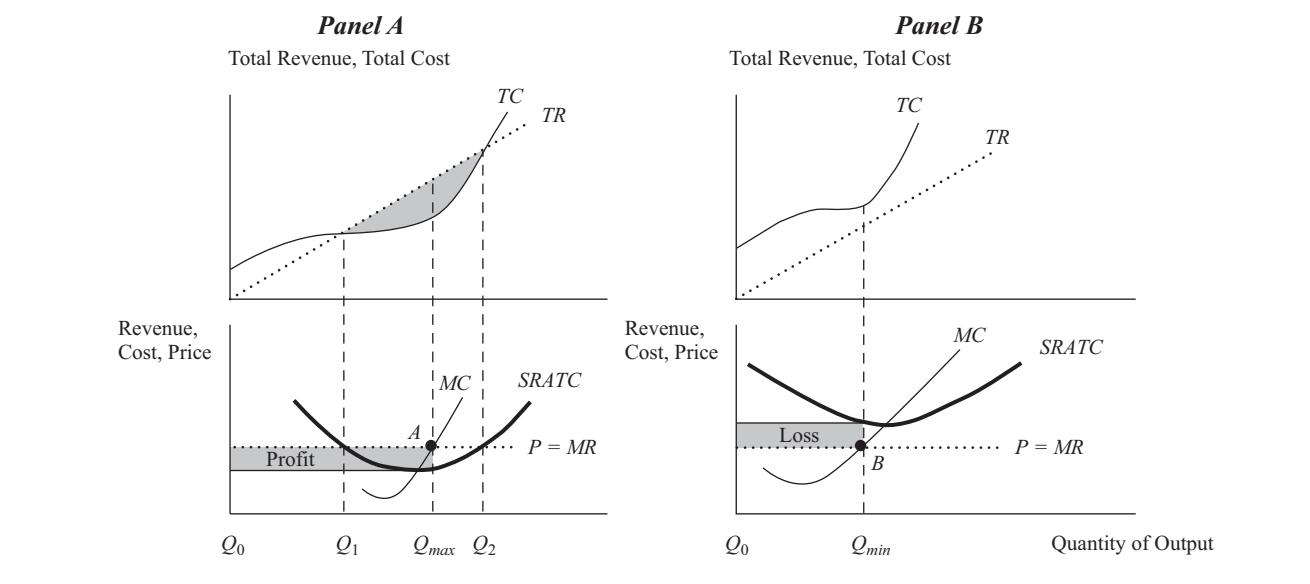


Exhibit 31 illustrates long-run profit maximization under perfect competition given the long-run average total cost curve when economies of scale occur. In the long run under perfect competition, the firm will operate at the minimum efficient scale point on its long-run average total cost curve. This least-cost point is illustrated in Exhibit 31 as point E at output level Q_E . In comparison to the point of minimum efficient scale, any other output quantity results in a higher cost.

In the short run given $SRATC_1$ and P_1 , the firm is making only normal profit because price equals average total cost at point A . By realizing economies of scale in the long run, the firm can move down the LRATC to $SRATC_2$ and produce Q_E . If the firm still receives P_1 , economic profit is forthcoming at Q_E in the amount of $(B - E)$ per unit. However, economic profit with no barriers to entry under perfect competition leads to more competitors, a greater market supply, and, subsequently, a lower price in the long run. The price to the firm will decline to P_2 , and economic profit will disappear with the long-run equilibrium for the firm occurring at point E , the minimum efficient scale. At point E , the firm is making only normal profit because in the long run under perfect competition, economic profit is zero.

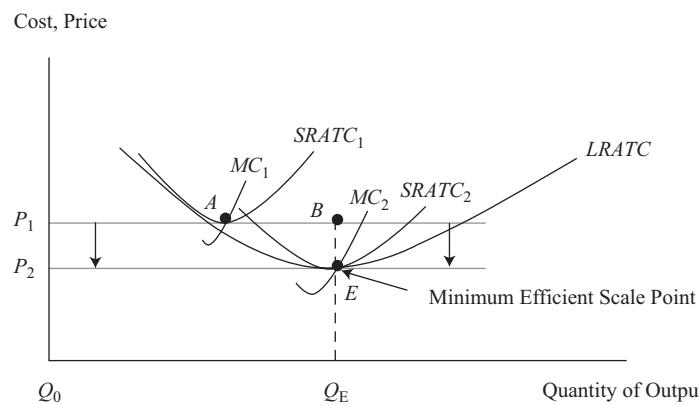
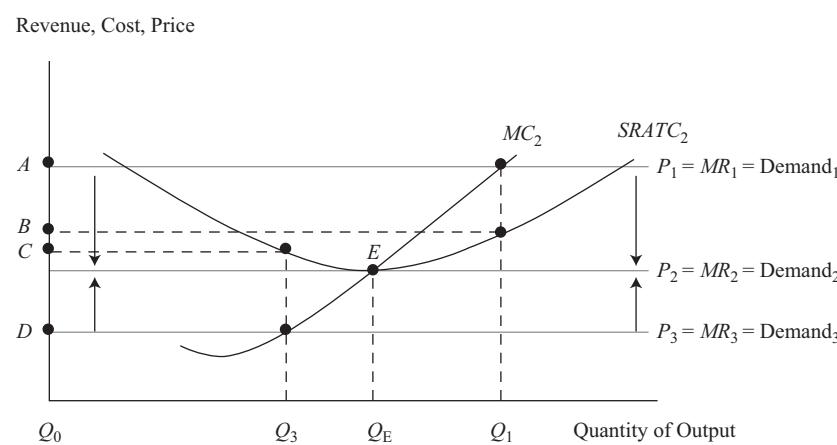
Exhibit 31 Long-Run Profit Maximization and Minimum Efficient Scale under Perfect Competition


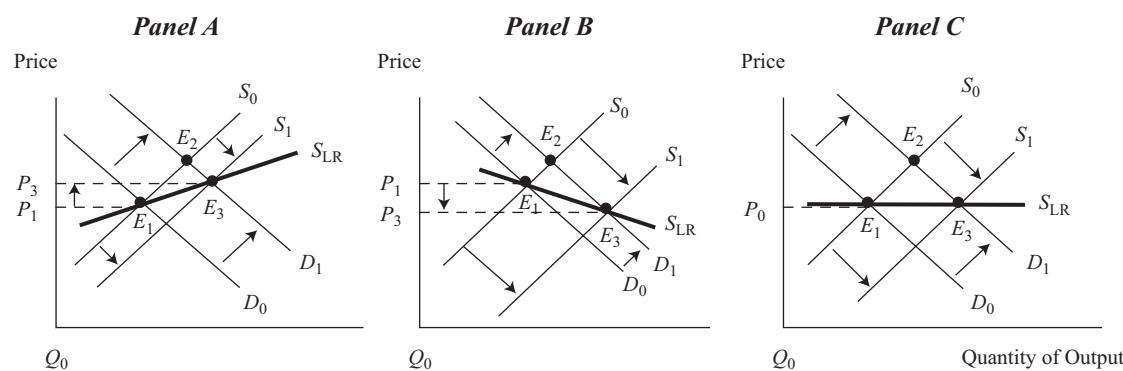
Exhibit 32 illustrates profit maximization and loss minimization in the long run when market prices change for a firm that is operating in a market of perfect competition at the minimum efficient scale point on its LRATC. ($SRATC_2$, MC_2 , Q_E , and point E are the same in both Exhibit 31 and Exhibit 32.) Although point E represents the lowest production cost, the quantity that maximizes profit or minimizes loss is determined where marginal revenue equals marginal cost. (Under perfect competition, price equals marginal revenue.) If price is at P_1 (which equals MR_1), the firm will produce Q_1 and accrue economic profit of $(A - B)$ per unit in the short run because price is greater than average total cost. In the long run, economic profit attracts new competitors who drive price down, resulting in zero economic profit. Profitability declines to the level at P_2 , where price is tangent to average total cost at point E . If price is at P_3 (which equals MR_3), the firm will produce Q_3 and realize an economic loss of $(C - D)$ per unit in the short run because average total cost is greater than price. In the long run, firms will exit the market; as a result, price rises to P_2 , eliminating economic losses. Again, profitability for the firm returns to the normal level at point E , where price matches average total cost. The long-term equilibrium for the firm occurs at point E , which corresponds to $Demand_2$, MR_2 , a price of P_2 , and an output level of Q_E .

Exhibit 32 Profit Maximization and Loss Minimization in the Long Run under Perfect Competition


3.1.7 The Long-Run Industry Supply Curve and What It Means for the Firm

The **long-run industry supply curve** shows the relationship between quantities supplied and output prices for an industry when firms are able to enter or exit the industry in response to the level of short-term economic profit (i.e., perfect competition) and when changes in industry output influence resource prices over the long run. Exhibit 33 illustrates three types of long-run supply curves based on increasing costs, decreasing costs, and constant costs to firms competing under perfect competition.

Exhibit 33 Long-Run Supply Curves for the Firm



An **increasing-cost industry** exists when prices and costs are higher when industry output is increased in the long run. This is demonstrated in Panel A. Assuming zero economic profit at E_1 , when demand increases from D_0 to D_1 , price rises and economic profit results at E_2 in the short run. Over the long term in response to this economic profit, new competitors will enter the industry and existing firms will expand output, resulting in an increase in supply from S_0 to S_1 and a long-term equilibrium at E_3 . If the increase in demand for resources from this output expansion leads to higher prices for some or all inputs, the industry as a whole will face higher production costs and charge a higher price for output. As indicated by S_{LR} in Panel A, the long-run supply curve for the industry will have a positive slope over the long run. The firm in an increasing-cost industry will experience higher resource costs so market price must rise in order to cover these costs. The petroleum, coal, and natural gas industries are prime examples of increasing-cost industries, where the supply response to long-run demand growth results in higher output prices because of the rising costs of energy production.

Panel B shows the case of a **decreasing-cost industry**, where the supply increase from S_0 to S_1 leads to a lower price for output in the market. Firms are able to charge a lower price because of a reduction in their resource costs. Decreasing costs can evolve from technological advances, producer efficiencies that come from a larger firm size, and economies of scale of resource suppliers (i.e., lower resource prices) that are passed on to resource buyers when industry output expands. The long-run supply curve for the industry will have a negative slope, as displayed by S_{LR} in Panel B. As a result, the firm in a decreasing-cost industry will experience lower resource costs and can then charge a lower price.⁴ Possible examples of decreasing-cost industries are semiconductors and personal computers, where the rapid growth in demand over the past decade has led to substantially lower prices.

⁴ The individual firm's supply curve is still upward sloping even though the industry's long-run supply curve is negatively sloped. This means that in the long run, the firm's supply curve shifts to the right when industry costs decrease. This results in a lower price charged by the firm for each quantity.

In some cases, firms in the industry will experience no change in resource costs and output prices over the long run. This type of industry is known as a **constant-cost industry**. This is displayed in Panel C, where the long-run supply curve (S_{LR}) for the industry is a horizontal line indicating a constant price level when the industry increases output.

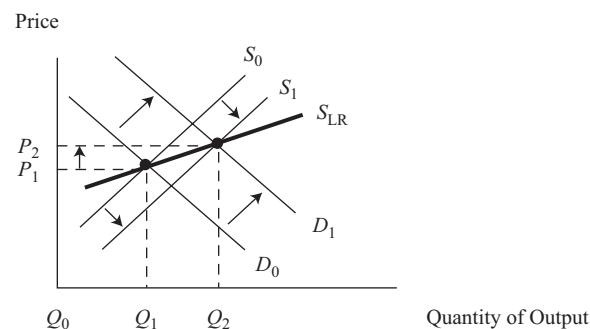
EXAMPLE 9

A Firm Operating in an Increasing-Cost Industry

Mirco Industries is a global manufacturer of outdoor recreational equipment in a market setting of easy entry and price competition. Company forecasts of total market demand for outdoor recreational products over the long run indicate robust growth in sales as families allocate more time to outdoor and leisure activities. This scenario looks promising for Mirco's earnings and shareholders' value. To assist Mirco in its assessment of this future scenario, industry analysts have presented Exhibit 34 to illustrate industry costs and the market supply curve over the long run.

Using Exhibit 34, what information would be of value to Mirco in identifying future production cost and price under a market growth scenario?

Exhibit 34



Note: Market demand increases from D_0 to D_1 . The market responds with an increase in supply from S_0 to S_1 . In the long run, the price of P_2 will be higher than the original price of P_1 .

Solution:

As indicated by the upward slope of the long-run industry supply curve, Mirco will experience an increase in production costs over the long run because of higher resource prices when the industry expands production and new firms enter the industry in response to an increase in market demand. To cover the higher production costs, Mirco will ultimately charge the higher market price of P_2 relative to the current price of P_1 .

3.2 Productivity

In general terms, **productivity** is defined as the average output per unit of input. Any production factor can be used as the input variable. However, it has been a common practice to use the labor resource as the basis for measuring productivity. In this regard, productivity is based on the number of workers used or the number of work-hours

performed. In many cases, labor is easier to quantify relative to the other types of resources used in production. As such, productivity is typically stated as output per worker or output per labor hour.

Why is productivity important? Cost minimization and profit maximization behavior dictate that the firm strives to maximize productivity—that is, produce the most output per unit of input or produce any given level of output with the least amount of inputs. A firm that lags behind the industry in productivity is at a competitive disadvantage and, as a result, is most likely to face decreases in future earnings and shareholders' wealth. An increase in productivity lowers production costs, which leads to greater profitability and investment value. Furthermore, productivity benefits (e.g., increased profitability) can be fully or partially distributed to other stakeholders of the business, such as consumers in the form of lower prices and employees in the form of enhanced compensation. Transferring some or all of the productivity rewards to non-equity holders creates synergies that benefit shareholders over time.

The benefits from increased productivity are as follows:

- Lower business costs, which translate into increased profitability.
- An increase in the market value of equity and shareholders' wealth resulting from an increase in profit.
- An increase in worker rewards, which motivates further productivity increases from labor.

Undoubtedly, increases in productivity reinforce and strengthen the competitive position of the firm over the long run. A fundamental analysis of a company should examine the firm's commitment to productivity enhancements and the degree to which productivity is integrated into the competitive nature of the industry or market. In some cases, productivity is not only an important promoter of growth in firm value over the long term, but it is also the key factor for economic survival. A business that lags the market in terms of productivity often finds itself less competitive, while at the same time confronting profit erosion and deterioration in shareholders' wealth. Whenever productivity is a consideration in the equity valuation of the firm, the first step for the analyst is to define measures of productivity. Typical productivity measures for the firm are based on the concepts of total product, average product, and marginal product of labor.

3.2.1 Total, Average, and Marginal Product of Labor

When measuring a firm's operating efficiency, it is easier and more practical to use a single resource factor as the input variable rather than a bundle of the different resources that the firm uses in producing units of output. As discussed in the previous section, labor is typically the input that is the most identifiable and calculable for measuring productivity. However, any input that is not difficult to quantify can be used. An example will illustrate the practicality of using a single factor input, such as labor, to evaluate the firm's output performance. A business that manually assembles widgets has 50 workers, one production facility, and an assortment of equipment and hand tools. The firm would like to assess its productivity when it utilizes these three types of input factors to produce widgets. In this case, the most appropriate method is to use labor as the input factor for determining productivity because the firm uses a variety of physical capital and only one plant building.

To illustrate the concepts of total product, average product, and marginal product, labor is used as the input variable. Exhibit 35 provides a summary of definitions and tabulations for these three concepts.

Exhibit 35 Definitions and Calculations for Total, Marginal, and Average Product of Labor

Term	Calculation
Total product	Sum of the output from all inputs during a time period; usually illustrated as the total output (TP or Q) using labor (L)
Average product	Total product divided by the quantity of a given input; measured as total product divided by the number of workers used at that output level; ($TP \div L$) or ($Q \div L$)
Marginal product	The amount of additional output resulting from using one more unit of input assuming other inputs are fixed; measured by taking the difference in total product and dividing by the change in the quantity of labor; ($\Delta TP \div \Delta L$) or ($\Delta Q \div \Delta L$)

Measured on the basis of the labor input, **total product** (TP or Q) is defined as the aggregate sum of production for the firm during a time period. As a measure of productivity, total product provides superficial information as to how effective and efficient the firm is in terms of producing output. For instance, three firms—Company A, Company B, and Company C—that comprise the entire industry have total output levels of 100,000 units, 180,000 units, and 200,000 units, respectively. Obviously, Company C dominates the market with a 41.7 percent share, followed by Company B's 37.5 percent share and Company A's 20.8 percent portion of the market. This information says little about how efficient each firm is in generating its total output level. Total product only provides an insight into the firm's production volume relative to the industry; it does not show how efficient the firm is in producing its output.

Average product (AP) measures the productivity of inputs on average and is calculated by dividing total product by the total number of units for a given input that is used to generate that output. Average product is usually measured on the basis of the labor input. It is a representative or overall measure of labor's productivity: Some workers are more productive than average, and others are less productive than average.

Given the aforementioned production levels for the three firms, Company A employs 100 workers, and Company B and Company C utilize 200 and 250 workers, respectively. Calculating average product of labor for each of the three firms yields the following productivity results: Company A→1,000 units of output per worker, Company B→900 units per worker, and Company C→800 units per worker. It is apparent that Company A is the most efficient firm, although it has the lowest share of the total market. Company C has the largest portion of the total market, but it is the least efficient of the three. Given that Company A can maintain its productivity advantage over the long run, it will be positioned to generate the greatest return on investment through lower costs and higher profit outcomes relative to the other firms in the market.

Marginal product (MP), also known as marginal return, measures the productivity of each unit of input and is calculated by taking the difference in total product from adding another unit of input (assuming other resource quantities are held constant). Typically, it is measured in terms of labor's performance; thereby, it is a gauge of productivity of the individual additional worker rather than an average across all workers.

Exhibit 36 provides a numerical illustration for total, average, and marginal products of labor.

Exhibit 36 Total, Average, and Marginal Product of Labor

Labor (L)	Total Product (TP_L)	Average Product (AP_L)	Marginal Product (MP_L)
0	0	—	—
1	100	100	100
2	210	105	110
3	300	100	90
4	360	90	60
5	400	80	40
6	420	70	20
7	350	50	(70)

Total product increases as the firm adds labor until worker 7, where at that point total production declines by 70 units. Obviously, the firm does not want to employ any worker that has negative productivity. In this case, no more than six workers are considered for employment with the firm.

At an employment level of five workers, AP and MP are 80 units ($400 \div 5$) and 40 units [$(400 - 360) \div (5 - 4)$], respectively. The productivity of the fifth worker is 40 units, while the average productivity for all five workers is 80 units, twice that of worker 5.

A firm has a choice of using total product, average product, marginal product, or some combination of the three to measure productivity. Total product does not provide an in-depth view of a firm's state of efficiency. It is simply an indication of a firm's output volume and potential market share. Therefore, average product and marginal product are better gauges of a firm's productivity because both can reveal competitive advantage through production efficiency. However, individual worker productivity is not easily measurable when workers perform tasks collectively. In this case, average product is the preferred measure of productivity performance.

3.2.2 Marginal Returns and Productivity

Referring to the marginal product column in Exhibit 36, worker 2 has a higher output of 110 units compared with worker 1 who produces 100 units; there is an increase in return when employees are added to the production process. This economic phenomenon is known as **increasing marginal returns**, where the marginal product of a resource increases as additional units of that input are employed. However, successive workers beyond number 2 have lower and lower marginal product to the point where the last worker has a negative return. This observation is called the **law of diminishing returns**. Diminishing returns can lead to a negative marginal product as evidenced with worker 7. There is no question that a firm does not want to employ a worker or input that has a negative impact on total output.

Initially, a firm can experience increasing returns from adding labor to the production process because of the concepts of specialization and division of labor. At first, by having too few workers relative to total physical capital, the understaffing situation requires employees to multi-task and share duties. As more workers are added, employees can specialize, become more adept at their individual functions, and realize an increase in marginal productivity. But after a certain output level, the law of diminishing returns becomes evident.

Assuming all workers are of equal quality and motivation, the decline in marginal product is related to the short run, where at least one resource (typically plant size, physical capital, and/or technology) is fixed. When more and more workers are added to a fixed plant size-technology-physical capital base, the marginal return of the labor

factor eventually decreases because the fixed input restricts the output potential of additional workers. One way of understanding the law of diminishing returns is to void the principle and assume that the concept of increasing returns lasts indefinitely. As more workers are added, or when any input is increased, the marginal output continuously increases. At some point, the world's food supply could be grown on one hectare of land or all new automobiles could be manufactured in one factory. Physically, the law of increasing returns is not possible in perpetuity, even though it can clearly be evident in the early stages of production.

Another element resulting in diminishing returns is the quality of labor itself. In the previous discussion, it was assumed all workers were of equal ability. However, that assumption may not be entirely valid when the firm's supply of labor has varying degrees of human capital. In that case, the business would want to employ the most productive workers first; then, as the firm's labor demand increased, less-productive workers would be hired. When the firm does not have access to an adequate supply of homogenous human capital, or for that matter any resource, diminishing marginal product occurs at some point.

EXAMPLE 10

Calculation and Interpretation of Total, Average, and Marginal Product

Average product and marginal product can be calculated on the basis of the production relationship between the number of machines and total product, as indicated in the first two columns of Exhibit 37.

- 1 Interpret the results for total, average, and marginal product.
- 2 Indicate where increasing marginal returns change to diminishing marginal returns.

Exhibit 37

Machines (K)	Total Product (TP_K)	Average Product (AP_K)	Marginal Product (MP_K)
0	0	—	—
1	1,000	1,000	1,000
2	2,500	1,250	1,500
3	4,500	1,500	2,000
4	6,400	1,600	1,900
5	7,400	1,480	1,000
6	7,500	1,250	100
7	7,000	1,000	(500)

Solution to 1:

Total product increases to six machines, where it tops out at 7,500. Because total product declines from machine 6 to machine 7, the marginal product for machine 7 is negative 500 units. Average product peaks at 1,600 units with four machines.

Solution to 2:

Increasing returns are evident up to machine 3, where marginal product equals 2,000 units of output. Beyond machine 3, decreasing returns develop because MP_K declines when more machines are added to the production process.

The data provided in Exhibit 38 show productivity changes for various US industries over the period 2000–2007. The coal mining and newspaper sectors have several years of negative changes in productivity, which do not reinforce prospects for long-term growth in profitability. Declines in productivity raise production costs and reduce profit. For the most part, the other industries have solid productivity increases from year to year, even though in some cases the change is volatile. On a trend basis, productivity increases appear to have peaked in the period 2002–2004 and then edged downward during the latter part of the period. Declining productivity makes a firm or industry less competitive over time; however, any adverse impact on profitability stemming from lower or negative productivity may be offset by rising demand for the product.

Exhibit 38 Productivity Changes for Selected Sectors, 2000–2007

Sector	NAICS¹	2000	2001	2002	2003	2004	2005	2006	2007
Coal mining	212,100	4.9%	-1.3%	-2.3%	1.5%	0.0%	-4.9%	-7.5%	1.2%
Newspaper publishers	511,110	5.5	-4.3	-0.6	5.1	-5.6	2.4	4.0	-1.8
Auto	336,100	-10.6	0.3	14.5	12.0	1.1	4.6	10.2	4.8
Commercial banking	522,110	3.9	-2.3	4.3	4.5	5.5	1.3	2.9	0.9
Merchandise stores	452,000	5.9	3.8	3.5	6.0	2.8	3.2	3.3	0.5
Air transportation	481,000	1.9	-5.3	9.9	10.2	12.7	7.6	5.1	1.8

¹North American Industry Classification System.

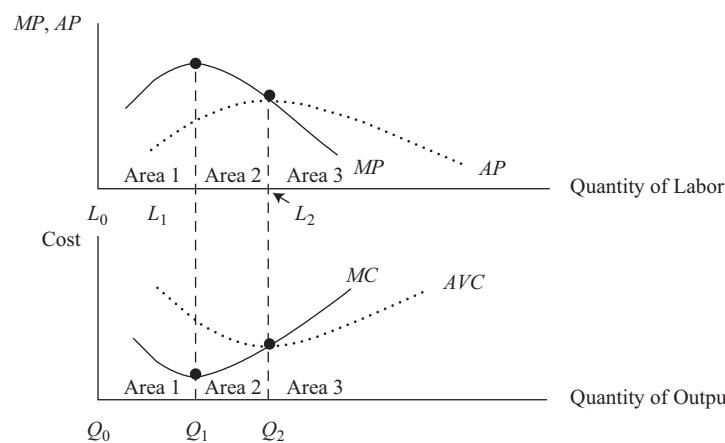
Note: Productivity is defined by the US Bureau of Labor Statistics as output per worker-hour.

Source: US Bureau of Labor Statistics, "Productivity and Costs by Industry: Annual Rates of Change."

Productivity is a key element in the determination of costs and profit to the firm, especially over the long term. Although productivity can fluctuate widely in the short run (as indicated in Exhibit 38) for a variety of reasons, secular patterns in output per unit of labor denote more meaningful relationships among productivity, costs, profits, and the competitive status of the firm with respect to the industry. To summarize, the analyst should study the productivity levels of the firm over the long run and do an evaluation of how the firm's efficiency compares with the industry standard. A firm that lags the industry in productivity may find itself at a competitive disadvantage with the end result of profit erosion and negative implications for shareholders' wealth. Once evident, productivity issues cause the firm's market value of equity to be discounted.

As previously discussed, a major determinant of the cost component of the profit equation is the degree of efficiency in which the firm uses resources in producing output as defined by the firm's production function. Given the relationship between output and inputs, marginal product (MP) and average product (AP) form the basis for marginal cost (MC) and average variable cost (AVC). Actually, MC and AVC are respective mirror images of MP and AP . Exhibit 39 illustrates this relationship in the short run by showing three areas of interest.

Exhibit 39 Relationship of Average Product and Marginal Product to Average Variable Cost and Marginal Cost in the Short Run



Area 1 shows an increasing MP from L_0 to L_1 . The increases in MP result in declining marginal costs from Q_0 to Q_1 . As MP or productivity peaks at L_1 , MC is minimized at Q_1 . Diminishing marginal returns take over in Areas 2 and 3, where a decreasing marginal product results in higher marginal costs. Not only does MP impact MC , but the shape of the AVC also is based on the pattern of AP . At L_2 , AP is maximized, while its corresponding output level of Q_2 is consistent with the minimum position on the AVC curve. Note that when MP is greater than AP , AP is increasing; when MP is less than AP , AP is declining. A similar relationship holds true for MC and AVC . When MC is less than AVC , AVC is decreasing; the opposite occurs when MC is greater than AVC . In Area 3, AP is declining, which creates an upturn in the AVC curve.

Technology, quality of human and physical capital, and managerial ability are key factors in determining the production function relationship between output and inputs. The firm's production function establishes what productivity is in terms of TP , MP , and AP . In turn, productivity significantly influences total, marginal, and average costs to the firm, and costs directly impact profit. Obviously, what happens at the production level in terms of productivity impacts the cost level and profitability.

Because revenue, costs, and profit are measured in monetary terms, the productivity of the different input factors requires comparison on a similar basis. In this regard, the firm wants to maximize output per monetary unit of input cost. This goal is denoted by the following expression:

$$\frac{MP_{input}}{P_{input}} \quad (7)$$

where MP_{input} is the marginal product of the input factor and P_{input} is the price of that factor (i.e., resource cost).

When using a combination of resources, a least-cost optimization formula is constructed as follows:

$$\frac{MP_1}{\text{Price of input 1}} = \dots = \frac{MP_n}{\text{Price of input } n} \quad (8)$$

where the firm utilizes n different resources. Using a two-factor production function consisting of labor and physical capital, Equation 9 best illustrates this rule of least cost:

$$\frac{MP_L}{P_L} = \frac{MP_K}{P_K} \quad (9)$$

where MP_L and MP_K are the marginal products of labor and physical capital, respectively. P_L is the price of labor or the wage rate, and P_K is the price of physical capital. For example, if MP_L/P_L equals two and MP_K/P_K is four, physical capital yields twice the output per monetary unit of input cost versus labor. It is obvious that the firm will want to use physical capital over labor in producing additional output because it provides more productivity on an equivalent cost basis. However, as more physical capital is employed, the firm's MP of capital declines because the law of diminishing returns impacts production. Physical capital is added until its ratio of MP per monetary unit of input cost matches that of labor: $MP_K/P_K = MP_L/P_L = 2$.⁵ At this point, both inputs are added when expanding output until their ratios differ. When their ratios diverge, the input with the higher ratio will be employed over the other lower ratio input when the firm increases production.

EXAMPLE 11

Determining the Optimal Input Combination

Canadian Global Electronic Corp. (CGEC) uses three types of labor—unskilled, semi-skilled, and skilled—in the production of electronic components. The firm's production technology allows for the substitution of one type of labor for another. Also, the firm buys labor in a perfectly competitive resource market in which the price of labor stays the same regardless of the number of workers hired. In the following table, the marginal productivity and compensation in Canadian dollars for each type of labor is displayed.

What labor type should the firm hire when expanding output?

Type of Labor	Marginal Product (MP_{input}) per Day	Compensation (P_{input}) per Day (\$)	$\frac{MP_{input}}{P_{input}}$
Unskilled (U)	200 units (MP_U)	100 (P_U)	2 units per \$
Semi-skilled (SS)	500 units (MP_{SS})	125 (P_{SS})	4 units per \$
Skilled (S)	1,000 units (MP_S)	200 (P_S)	5 units per \$

Solution:

The firm minimizes cost and enhances profitability by adding skilled labor over the other two types because it has the highest ratio of MP to input price. As the marginal product of skilled labor declines with additional workers, MP_S/P_S decreases. When it declines to the same value as semi-skilled labor, both skilled and semi-skilled workers are added because their productivity per Canadian dollar of input cost is identical. Again, a diminishing marginal product decreases both ratios. When all three labor inputs have the same MP_{input}/P_{input} , the firm will add all three labor types at the same time when expanding output.

⁵ This assumes that MP_L is independent of physical capital K . However, as more K is used, MP_L could actually increase because labor will become more productive when using more physical capital. In this case, $MP_K/P_K = MP_L/P_L$ at some point between 2 and 4.

Equations 7, 8, and 9 derive the physical output per monetary unit of input cost. However, to determine the profit-maximizing utilization level of an input, the firm must measure the revenue value of the input's *MP* and then compare this figure with the cost of the input. The following equations represent this relationship:

$$\text{Marginal product} \times \text{Product price} = \text{Price of the input} \quad (10)$$

$$\text{Marginal revenue product} = \text{Price of the input} \quad (11)$$

Marginal revenue product (MRP) is calculated as the *MP* of an input unit times the price of the product. This term measures the value of the input to the firm in terms of what the input contributes to *TR*. It is also defined as the change in *TR* divided by the change in the quantity of the resource employed. If an input's *MRP* exceeds its cost, a contribution to profit is evident. For example, when the *MP* of the last unit of labor employed is 100 and the product price is 2.00, the *MRP* for that unit of labor (MRP_L) is 200. When the input price of labor is 125, the surplus value or contribution to profit is 75. In contrast, if *MRP* is less than the input's price, a loss would be incurred from employing that input unit. If the *MP* of the next unit of labor is 50 with a product price of 2.00, MRP_L will now be 100. With the same labor cost of 125, the firm would incur a loss of 25 when employing this input unit. Profit maximization occurs when the *MRP* equates to the price or cost of the input for each type of resource that is used in the production process.

In the case of multiple factor usage, the following equation holds true for n inputs:

$$\frac{MRP_1}{\text{Price of input } 1} = \dots = \frac{MRP_n}{\text{Price of input } n} = 1 \quad (12)$$

When profit is maximized, *MRP* equals the input price for each type of resource used and all MRP_{input}/P_{input} are equal to one.

EXAMPLE 12

Profit Maximization Using the Marginal Revenue Product and Resource Cost Approach

Using the data from the previous case of Canadian Global Electronic Corp., the table below shows the *MRP* per labor type when product price in Canadian dollars is \$0.50. *MRP* per day is calculated as the *MP* per type of labor from Example 11 multiplied by the product price.

Which type of labor contributes the most to profitability?

Type of Labor	Marginal Revenue Product (MRP_{input}) per Day (\$)	Compensation (P_{input}) per Day (\$)	$\frac{MRP_{input}}{P_{input}}$
Unskilled (U)	100 (MRP_U)	100 (P_U)	1.0
Semi-skilled (SS)	250 (MRP_{SS})	125 (P_{SS})	2.0
Skilled (S)	500 (MRP_S)	200 (P_S)	2.5

Solution:

Calculating the MRP_{input}/P_{input} values for the different labor categories yields ratio numbers of 1.0, 2.0, and 2.5 for unskilled, semi-skilled, and skilled labor, respectively. The firm adds skilled labor first because it is the most profitable to employ, as indicated by MRP_S/P_S being the highest ratio of the three labor inputs. The contribution to profit by employing the next skilled worker is \$300, calculated as $(\$500 - \$200)$. However, with the employment of additional skilled

workers, MRP_S declines because of diminishing returns that are associated with the MP component. At the point where the skilled labor ratio drops below 2.0—for example, to 1.5—semi-skilled labor becomes feasible to hire because its MRP exceeds its compensation by more than that of skilled labor.⁶ Again, the diminishing returns effect decreases MRP when additional semi-skilled workers are hired. In the case of unskilled labor, MRP_U equals the cost of labor; hence, no further contribution to profit accrues from adding this type of labor. In fact, adding another unskilled worker would probably reduce total profit because the next worker's compensation is likely to exceed MRP as a result of a declining MP . The input level that maximizes profit is where $MRP_U/P_U = MRP_{SS}/P_{SS} = MRP_S/P_S = 1$.

SUMMARY

When assessing financial performance, a micro exploration of a firm's profitability reveals more information to the analyst relative to the typical macro examination of overall earnings. Crucial issues evolve when the firm fails to reward investors properly for their equity commitment and when the firm's operating status is non-optimal in regard to resource employment, cost minimization, and profit maximization.

Among the points made in this reading are the following:

- The two major concepts of profits are accounting profit and economic profit. Economic profit equals accounting profit minus implicit opportunity costs not included in accounting costs. Profit in the theory of the firm refers to economic profit.
- Normal profit is an economic profit of zero. A firm earning a normal profit is earning just enough to cover the explicit and implicit costs of resources used in running the firm, including, most importantly for publicly traded corporations, debt and equity capital.
- Economic profit is a residual value in excess of normal profit and results from access to positive NPV investment opportunities.
- The factors of production are the inputs to the production of goods and services and include land, labor, capital, and materials.
- Profit maximization occurs at the following points:
 - Where the difference between total revenue and total costs is the greatest.
 - Where marginal revenue equals marginal cost.
 - Where marginal revenue product equals the resource cost for each type of input.
- When total costs exceed total revenue, loss minimization occurs where the difference between total costs and total revenue is the least.
- In the long run, all inputs to the firm are variable, which expands profit potential and the number of cost structures available to the firm.

⁶ The next semi-skilled worker contributes \$125 (derived as \$250 – \$125) per day to profit, while the next skilled worker's contribution, based on a ratio of 1.5, is \$100 (MRP_S of \$300 minus compensation of \$200 per day).

- Under perfect competition, long-run profit maximization occurs at the minimum point of the firm's long-run average total cost curve.
- In an economic loss situation, a firm can operate in the short run if total revenue covers variable cost but is inadequate to cover fixed cost; however, in the long run, the firm will exit the market if fixed costs are not covered in full.
- In an economic loss situation, a firm shuts down in the short run if total revenue does not cover variable cost in full and eventually exits the market if the shortfall is not reversed.
- Economies of scale lead to lower average total cost; diseconomies of scale lead to higher average total cost.
- A firm's production function defines the relationship between total product and inputs.
- Average product and marginal product, which are derived from total product, are key measures of a firm's productivity.
- Increases in productivity reduce business costs and enhance profitability.
- An industry supply curve that is positively sloped in the long run will increase production costs to the firm. An industry supply curve that is negatively sloped in the long run will decrease production costs to the firm.
- In the short run, assuming constant resource prices, increasing marginal returns reduce the marginal costs of production and decreasing marginal returns increase the marginal costs of production.

PRACTICE PROBLEMS

- 1 Normal profit is *best* described as:
 - A zero economic profit.
 - B total revenue minus all explicit costs.
 - C the sum of accounting profit plus economic profit.
- 2 A firm supplying a commodity product in the marketplace is *most likely* to receive economic rent if:
 - A demand increases for the commodity and supply is elastic.
 - B demand increases for the commodity and supply is inelastic.
 - C supply increases for the commodity and demand is inelastic.
- 3 Entrepreneurs are *most likely* to receive payment or compensation in the form of:
 - A rent.
 - B profit.
 - C wages.
- 4 The marketing director for a Swiss specialty equipment manufacturer estimates the firm can sell 200 units and earn total revenue of CHF500,000. However, if 250 units are sold, revenue will total CHF600,000. The marginal revenue per unit associated with marketing 250 units instead of 200 units is *closest* to:
 - A CHF 2,000.
 - B CHF 2,400.
 - C CHF 2,500.
- 5 An agricultural firm operating in a perfectly competitive market supplies wheat to manufacturers of consumer food products and animal feeds. If the firm were able to expand its production and unit sales by 10% the *most likely* result would be:
 - A a 10% increase in total revenue.
 - B a 10% increase in average revenue.
 - C an increase in total revenue of less than 10%.
- 6 An operator of a ski resort is considering offering price reductions on weekday ski passes. At the normal price of €50 per day, 300 customers are expected to buy passes each weekday. At a discounted price of €40 per day 450 customers are expected to buy passes each weekday. The marginal revenue per customer earned from offering the discounted price is *closest* to:
 - A €20.
 - B €40.
 - C €50.
- 7 The marginal revenue per unit sold for a firm doing business under conditions of perfect competition will *most likely* be:
 - A equal to average revenue.
 - B less than average revenue.
 - C greater than average revenue.

The following information relates to Questions 8–10

A firm's director of operations gathers the following information about the firm's cost structure at different levels of output:

Exhibit 1

Quantity (Q)	Total Fixed Cost (TFC)	Total Variable Cost (TVC)
0	200	0
1	200	100
2	200	150
3	200	200
4	200	240
5	200	320

- 8 Refer to the data in Exhibit 1. When quantity produced is equal to 4 units, the average fixed cost (*AFC*) is *closest* to:
- A 50.
 - B 60.
 - C 110.
- 9 Refer to the data in Exhibit 1. When the firm increases production from 4 to 5 units, the marginal cost (*MC*) is *closest* to:
- A 40.
 - B 64.
 - C 80.
- 10 Refer to the data in Exhibit 1. The level of unit production resulting in the lowest average total cost (*ATC*) is *closest* to:
- A 3.
 - B 4.
 - C 5.
-
- 11 The short-term breakeven point of production for a firm operating under perfect competition will *most likely* occur when:
- A price is equal to average total cost.
 - B marginal revenue is equal to marginal cost.
 - C marginal revenue is equal to average variable costs.
- 12 The short-term shutdown point of production for a firm operating under perfect competition will *most likely* occur when:
- A price is equal to average total cost.
 - B marginal revenue is equal to marginal cost.
 - C marginal revenue is less than average variable costs.

- 13 When total revenue is greater than total variable costs but less than total costs, in the short term a firm will *most likely*:
- A exit the market.
 - B stay in the market.
 - C shut down production.
- 14 A profit maximum is *least likely* to occur when:
- A average total cost is minimized.
 - B marginal revenue equals marginal cost.
 - C the difference between total revenue and total cost is maximized.
- 15 A firm that increases its quantity produced without any change in per-unit cost is experiencing:
- A economies of scale.
 - B diseconomies of scale.
 - C constant returns to scale.
- 16 A firm is operating beyond minimum efficient scale in a perfectly competitive industry. To maintain long-term viability the *most likely* course of action for the firm is to:
- A operate at the current level of production.
 - B increase its level of production to gain economies of scale.
 - C decrease its level of production to the minimum point on the long-run average total cost curve.
- 17 Under conditions of perfect competition, in the long run firms will *most likely* earn:
- A normal profits.
 - B positive economic profits.
 - C negative economic profits.
- 18 A firm engages in the development and extraction of oil and gas, the supply of which is price inelastic. The *most likely* equilibrium response in the long run to an increase in the demand for petroleum is that oil prices:
- A increase, and extraction costs per barrel fall.
 - B increase, and extraction costs per barrel rise.
 - C remain constant, and extraction costs per barrel remain constant.
- 19 A firm develops and markets consumer electronic devices in a perfectly competitive, decreasing-cost industry. The firm's products have grown in popularity. The *most likely* equilibrium response in the long run to rising demand for such devices is for selling prices to:
- A fall and per-unit production costs to decrease.
 - B rise and per-unit production costs to decrease.
 - C remain constant and per-unit production costs to remain constant.

The following information relates to Questions 20–21

The manager of a small manufacturing firm gathers the following information about the firm's labor utilization and production:

Exhibit 2

Labor (<i>L</i>)	Total Product (<i>TP</i>)
0	0
1	150
2	320
3	510
4	660
5	800

- 20** Refer to the data in Exhibit 2. The number of workers resulting in the highest level of average product of labor is *closest* to:
- A 3.
 - B 4.
 - C 5.
- 21** Refer to the data in Exhibit 2. The marginal product of labor demonstrates increasing returns for the firm if the number of workers is *closest* to but not more than:
- A 2.
 - B 3.
 - C 4.
-
- 22** A firm experiencing an increase in the marginal product of labor employed would *most likely*:
- A allow an increased number of workers to specialize and become more adept at their individual functions.
 - B find that an increase in workers cannot be efficiently matched by other inputs that are fixed such as property, plant, and equipment.
 - C find that the supply of skilled workers is limited, and additional workers lack essential skills and aptitudes possessed by the current workforce.
- 23** For a manufacturing company to achieve the most efficient combination of labor and capital, and therefore minimize total costs for a desired level of output, it will *most likely* attempt to equalize the:
- A average product of labor to the average product of capital.
 - B marginal product per unit of labor to the marginal product per unit of capital.
 - C marginal product obtained per dollar spent on labor to the marginal product per dollar spent on capital.
- 24** A firm will expand production by 200 units and must hire at least one additional worker. The marginal product per day for one additional unskilled worker is 100 units. The marginal product per day for one additional skilled worker is 200 units. Wages per day are \$200 for an unskilled worker and \$450 for a skilled worker. The firm will *most likely* minimize costs at the higher level of production by hiring:
- A one additional skilled worker.

- B two additional unskilled workers.
- C either a skilled worker or two unskilled workers.
- 25 A Mexican firm employs unskilled, semi-skilled, and skilled labor in a cost-minimizing mix at its manufacturing plant. The marginal product of unskilled labor is considerably lower than semi-skilled and skilled labor, but the equilibrium wage for unskilled labor is only 300 pesos per day. The government passes a law that mandates a minimum wage of 400 pesos per day. Equilibrium wages for semi-skilled and skilled labor exceed this minimum wage and therefore are not affected by the new law. The firm will *most likely* respond to the imposition of the minimum wage law by:
- A employing more unskilled workers at its plant.
- B employing fewer unskilled workers at its plant.
- C keeping the mix of unskilled, semi-skilled, and skilled workers the same.

The following information relates to Questions

26–27

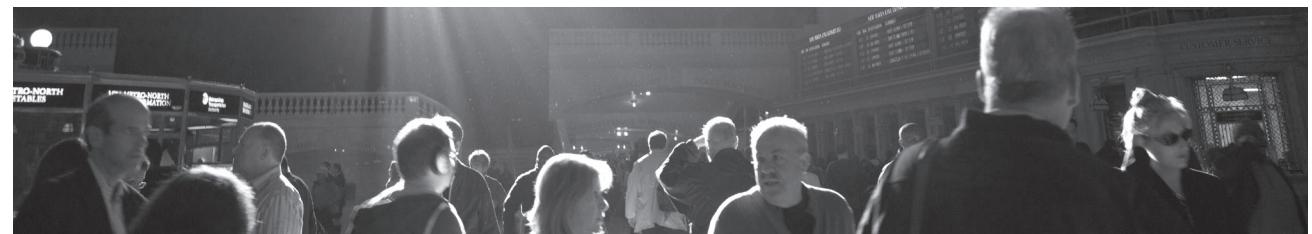
A firm produces handcrafted wooden chairs, employing both skilled craftsmen and automated equipment in its plant. The selling price of a chair is €100. A craftsman earns €900 per week and can produce ten chairs per week. Automated equipment leased for €800 per week can produce ten chairs per week.

- 26 The marginal revenue product (per week) of hiring an additional craftsman is *closest* to:
- A €100.
- B €900.
- C €1,000.
- 27 The firm would like to increase weekly output by 50 chairs. The firm would *most likely* enhance profits by:
- A hiring additional craftsmen.
- B leasing additional automated equipment.
- C leasing additional automated equipment and hiring additional craftsmen in equal proportion.
-

SOLUTIONS

- 1 A is correct. Normal profit is the level of accounting profit such that implicit opportunity costs are just covered; thus, it is equal to a level of accounting profit such that economic profit is zero.
- 2 B is correct. Economic rent results when a commodity is fixed in supply (highly inelastic) and the market price is higher than what is required to bring the commodity to market. An increase in demand in this circumstance would result in a rising price and increased potential for economic rent.
- 3 B is correct. Profit is the return to entrepreneurship for its contribution to the economic process.
- 4 A is correct. Marginal revenue per unit is defined as the change in total revenue divided by the change in quantity sold. $MR = \Delta TR \div \Delta Q$. In this case, change in total revenue equals CHF100,000, and change in total units sold equals 50. $CHF100,000 \div 50 = CHF2,000$.
- 5 A is correct. In a perfectly competitive market, an increase in supply by a single firm will not affect price. Therefore, an increase in units sold by the firm will be matched proportionately by an increase in revenue.
- 6 A is correct. Marginal revenue per unit is defined as the change in total revenues divided by the change in quantity sold. $MR = \Delta TR \div \Delta Q$. In this case, change in total revenue per day equals €3,000 [(450 x €40) – (300 x €50)], and change in units sold equals 150 (450 – 300). $€3,000 \div 150 = €20$.
- 7 A is correct. Under perfect competition, a firm is a price taker at any quantity supplied to the market, and $AR = MR = \text{Price}$.
- 8 A is correct. Average fixed cost is equal to total fixed cost divided by quantity produced: $AFC = TFC/Q = 200/4 = 50$.
- 9 C is correct. Marginal cost is equal to the change in total cost divided by the change in quantity produced. $MC = \Delta TC/\Delta Q = 80/1 = 80$.
- 10 C is correct. Average total cost is equal to total cost divided by quantity produced. At 5 units produced the average total cost is 104. $ATC = TC/Q = 520/5 = 104$.
- 11 A is correct. Under perfect competition, price equals marginal revenue. A firm breaks even when marginal revenue equals average total cost.
- 12 C is correct. The firm should shut down production when marginal revenue is less than average variable cost.
- 13 B is correct. When total revenue is enough to cover variable costs but not total fixed costs in full, the firm can survive in the short run but would be unable to maintain financial solvency in the long run.
- 14 A is correct. The quantity at which average total cost is minimized does not necessarily correspond to a profit maximum.
- 15 C is correct. Output increases in the same proportion as input increases occur at constant returns to scale.
- 16 C is correct. The firm operating at greater than long-run efficient scale is subject to diseconomies of scale. It should plan to decrease its level of production.
- 17 A is correct. Competition should drive prices down to long-run marginal cost, resulting in only normal profits being earned.

- 18** B is correct. The development and extraction of scarce oil and gas is an increasing-cost industry. A positive shift in demand will cause firms to increase supply, but at higher costs. The higher costs associated with increasing supply will cause prices to rise.
- 19** A is correct. A positive shift in demand will cause firms to increase supply, but at decreasing costs. The decreasing cost per unit will be passed on to consumers and cause prices to fall in the long run.
- 20** A is correct. Three workers produce the highest average product equal to 170. $AP = 510/3 = 170$.
- 21** B is correct. Marginal product is equal to the change in total product divided by the change in labor. The increase in MP from 2 to 3 workers is 190: $MP = \Delta TP/\Delta L = (510 - 320)/(3 - 2) = 190/1 = 190$.
- 22** A is correct. Adding new workers in numbers sufficient for them to specialize in their roles and functions should increase marginal product of labor.
- 23** C is correct. Costs are minimized when substitution of labor for capital (or the reverse) does not result in any cost savings, which is the case when the marginal product per dollar spent is equalized across inputs.
- 24** B is correct. An expansion in production by 200 units can be achieved by two unskilled workers at a total cost of \$400, or \$2 per unit produced. $\$400/200 = \2 per unit produced.
- 25** B is correct. The firm employs labor of various types in a cost-minimizing combination. Profit is maximized when marginal revenue product is equalized across each type of labor input. If the wage rate of unskilled workers increases, the marginal product produced per dollar spent to employ unskilled labor will decline. The original employment mix is no longer optimal, so the firm will respond by shifting away from unskilled workers to workers whose wages are unaffected by the minimum wage law.
- 26** C is correct. The marginal revenue product is the marginal product of an additional craftsman (10 chairs) times the price per chair (€100). $10 \times €100 = €1,000$.
- 27** B is correct. The marginal revenue product for additional power tools is €1,000, which exceeds the €800 cost of the tools by €200. $(10 \times €100 = €1,000) - €800 = €200$.



Glossary

Abnormal profit Equal to accounting profit less the implicit opportunity costs not included in total accounting costs; the difference between total revenue (TR) and total cost (TC).

Accounting (or explicit) costs Payments to non-owner parties for services or resources they supply to the firm.

Accounting loss When accounting profit is negative.

Accounting profit Income as reported on the income statement, in accordance with prevailing accounting standards, before the provisions for income tax expense. Also called *income before taxes or pretax income*.

Average fixed cost Total fixed cost divided by quantity.

Average product Measures the productivity of inputs on average and is calculated by dividing total product by the total number of units for a given input that is used to generate that output.

Average revenue Quantity sold divided into total revenue.

Average total cost Total costs divided by quantity.

Average variable cost Total variable cost divided by quantity.

Break-even point The number of units produced and sold at which the company's net income is zero (revenues = total costs); in the case of perfect competition, the quantity where price, average revenue, and marginal revenue equal average total cost.

Constant returns to scale The characteristic of constant per-unit costs in the presence of increased production.

Constant-cost industry When firms in the industry experience no change in resource costs and output prices over the long run.

Decreasing returns to scale Increase in cost per unit resulting from increased production.

Decreasing-cost industry An industry in which per-unit costs and output prices are lower when industry output is increased in the long run.

Diseconomies of scale Increase in cost per unit resulting from increased production.

Economic costs All the remuneration needed to keep a productive resource in its current employment or to acquire the resource for productive use; the sum of total accounting costs and implicit opportunity costs.

Economic loss The amount by which accounting profit is less than normal profit.

Economic profit Equal to accounting profit less the implicit opportunity costs not included in total accounting costs; the difference between total revenue (TR) and total cost (TC). Also called *abnormal profit or supernormal profit*.

Economic rent The surplus value that results when a particular resource or good is fixed in supply and market price is higher than what is required to bring the resource or good onto the market and sustain its use.

Economies of scale Reduction in cost per unit resulting from increased production.

Elasticity of supply A measure of the sensitivity of quantity supplied to a change in price.

Imperfect competition A market structure in which an individual firm has enough share of the market (or can control a certain segment of the market) such that it is able to exert some influence over price.

Increasing marginal returns Where the marginal product of a resource increases as additional units of that input are employed.

Increasing returns to scale Reduction in cost per unit resulting from increased production.

Increasing-cost industry An industry in which per-unit costs and output prices are higher when industry output is increased in the long run.

Inelastic supply Said of supply that is insensitive to the price of goods sold.

Inelastic Insensitive to price changes.

Law of diminishing returns The smallest output that a firm can produce such that its long run average costs are minimized.

Long-run average total cost curve The curve describing average total costs when no costs are considered fixed.

Long-run industry supply curve A curve describing the relationship between quantity supplied and output prices when no costs are considered fixed.

Marginal cost The cost of producing an additional unit of a good.

Marginal product Measures the productivity of each unit of input and is calculated by taking the difference in total product from adding another unit of input (assuming other resource quantities are held constant).

Marginal revenue product The amount of additional revenue received from employing an additional unit of an input.

Marginal revenue The change in total revenue divided by the change in quantity sold; simply, the additional revenue from selling one more unit.

Market structure The competitive environment (perfect competition, monopolistic competition, oligopoly, and monopoly).

Minimum efficient scale The smallest output that a firm can produce such that its long run average cost is minimized.

Monopolist Said of an entity that is the only seller in its market.

Normal profit The level of accounting profit needed to just cover the implicit opportunity costs ignored in accounting costs.

Perfect competition A market structure in which the individual firm has virtually no impact on market price, because it is assumed to be a very small seller among a very large number of firms selling essentially identical products.

Planning horizon A time period in which all factors of production are variable, including technology, physical capital, and plant size.

Price takers Producers that must accept whatever price the market dictates.

Price The market price as established by the interactions of the market demand and supply factors.

Production function Provides the quantitative link between the level of output that the economy can produce and the inputs used in the production process.

Productivity The amount of output produced by workers in a given period of time—for example, output per hour worked; measures the efficiency of labor.

Quantity demanded The amount of a product that consumers are willing and able to buy at each price level.

Quantity The amount of a product that consumers are willing and able to buy at each price level.

Quasi-fixed cost A cost that stays the same over a range of production but can change to another constant level when production moves outside of that range.

Shareholder wealth maximization To maximize the market value of shareholders' equity.

Short-run average total cost curve The curve describing average total costs when some costs are considered fixed.

Short-run supply curve The section of the marginal cost curve that lies above the minimum point on the average variable cost curve.

Shutdown point The point at which average revenue is less than average variable cost.

Supernormal profit Equal to accounting profit less the implicit opportunity costs not included in total accounting costs; the difference between total revenue (TR) and total cost (TC).

Theory of the consumer The branch of microeconomics that deals with consumption—the demand for goods and services—by utility-maximizing individuals.

Theory of the firm The branch of microeconomics that deals with the supply of goods and services by profit-maximizing firms.

Total costs The summation of all costs, where costs are classified according to fixed or variable.

Total fixed cost The summation of all expenses that do not change when production varies.

Total product The aggregate sum of production for the firm during a time period.

Total revenue Price times the quantity of units sold.

Total variable cost The summation of all variable expenses.